Universidad de Alcalá
COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

ACTA DE EVALUACIÓN DE LA TESIS DOCTORAL

(FOR EVALUATION OF THE ACT DOCTORAL THESIS)

Año académico (*academic year):* 2016/17

DOCTORANDO *(candidate PHD)*: **FERNÁNDEZ LÓPEZ, CARLOS**
D.N.I./PASAPORTE (*Id.Passport*): ******8344B**
PROGRAMA DE DOCTORADO *(Academic Committee of the Programme)*: **D347-TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES**
DEPARTAMENTO DE *(Department)*: **AUTOMÁTICA**
TITULACIÓN DE DOCTOR EN *(Phd title)*: **DOCTOR/A POR LA UNIVERSIDAD DE ALCALÁ**

En el día de hoy 23/09/16, reunido el tribunal de evaluación, constituido por los miembros que suscriben el presente Acta, el aspirante defendió su Tesis Doctoral **con Mención Internacional** (*In today assessment met the court, consisting of the members who signed this Act, the candidate defended his doctoral thesis with mention as International Doctorate*), elaborada bajo la dirección de (*prepared under the direction of*) MIGUEL ANGEL SOTELO VAZQUEZ // DAVID FERNÁNDEZ LLORCA.

Sobre el siguiente tema *(Title of the doctoral thesis)*: *ROAD SCENE INTERPRETATION FOR AUTONOMOUS NAVIGATION FUSING STEREO VISION AND DIGITAL MAPS*

Finalizada la defensa y discusión de la tesis, el tribunal acordó otorgar la CALIFICACIÓN GLOBAL[1] de (**no apto, aprobado, notable** y **sobresaliente**) *(After the defense and defense of the thesis, the court agreed to grant the GLOBAL RATING (fail, pass, good and excellent)*: EXCELLENT (SOBRESALIENTE)

Alcalá de Henares, a 23 de 9 de 2016

Fdo. *(Signed)*: Sebastián Sánchez

Fdo. *(Signed)*: IGNACIO PARRA ALONSO

Fdo. *(Signed)*:

Fdo. *(Signed)*:

Fdo. *(Signed)*:

FIRMA DEL ALUMNO (candidate's signature),

Fdo. *(Signed)*: CARLOS FDEZ LOPEZ

Con fecha 27 de septiembre de 2016 la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado, a la vista de los votos emitidos de manera anónima por el tribunal que ha juzgado la tesis, resuelve:

☒ Conceder la Mención de "Cum Laude"
☐ No conceder la Mención de "Cum Laude"

La Secretaria de la Comisión Delegada

[1] La calificación podrá ser **"no apto" "aprobado" "notable" y "sobresaliente"**. El tribunal podrá otorgar la mención de "cum laude" si la calificación global es de sobresaliente y se emite en tal sentido el voto secreto positivo por unanimidad. (*The grade may be **"fail" "pass" "good"** or **"excellent"**. The panel may confer the distinction of "cum laude" if the overall grade is "Excellent" and has been awarded unanimously as such after secret voting.).*

INCIDENCIAS / OBSERVACIONES:
(Incidents / Comments)

Universidad de Alcalá

COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

En aplicación del art. 14.7 del RD. 99/2011 y el art. 14 del Reglamento de Elaboración, Autorización y Defensa de la Tesis Doctoral, la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado y Doctorado, en sesión pública de fecha 27 de septiembre, procedió al escrutinio de los votos emitidos por los miembros del tribunal de la tesis defendida por *FERNÁNDEZ LÓPEZ, CARLOS*, el día 23 de septiembre de 2016, titulada *ROAD SCENE INTERPRETATION FOR AUTONOMOUS NAVIGATION FUSING STEREO VISION AND DIGITAL MAPS*, para determinar, si a la misma, se le concede la mención "cum laude", arrojando como resultado el voto favorable de todos los miembros del tribunal.

Por lo tanto, la Comisión de Estudios Oficiales de Posgrado **resuelve otorgar** a dicha tesis la

MENCIÓN "CUM LAUDE"

Alcalá de Henares, 28 de septiembre de 2016
EL PRESIDENTE DE LA COMISIÓN DE ESTUDIOS
OFICIALES DE POSGRADO Y DOCTORADO

UNIVERSIDAD DE ALCALÁ · COMISIÓN DE ESTUDIOS OFICIALES DE POSGRADO Y DOCTORADO ·

Juan Ramón Velasco Pérez

Copia por e-mail a:
Doctorando: FERNÁNDEZ LÓPEZ, CARLOS
Secretario del Tribunal: IGNACIO PARRA ALONSO
Directores de Tesis: MIGUEL ANGEL SOTELO VAZQUEZ // DAVID FERNÁNDEZ LLORCA

Universidad de Alcalá

Campus Universitario
Dpto. de Automática
Ctra. Madrid-Barcelona, Km. 36.6
28805 Alcalá de Henares (Madrid)
Telf: +34 91 885 65 94

D. MIGUEL ÁNGEL SOTELO VÁZQUEZ y D. DAVID FERNÁNDEZ LLORCA, Profesor Catedrático de Universidad y Profesor Titular de Universidad respectivamente del Área de Conocimiento de Ingeniería de Sistemas y Automática de la Universidad de Alcalá,

CERTIFICAN

Que la tesis "**Road scene interpretation for autonomous navigation fusing stereo vision and digital map**", presentada por D. Carlos Fernández López, realizada en el Departamento de Automática bajo nuestra dirección, reúne méritos suficientes para optar al grado de Doctor, por lo que puede procederse a su depósito y lectura.

Alcalá de Henares, June 30, 2016.

Fdo.: Dr. D. Miguel Ángel Sotelo Vázquez Fdo.: Dr. D. David Fernández Llorca

Universidad de Alcalá

Campus Universitario
Dpto. de Automática
Ctra. Madrid-Barcelona, Km. 36.6
28805 Alcalá de Henares (Madrid)
Telf: +34 91 885 65 94

D. Carlos Fernández López ha realizado en el Departamento de automática y bajo la direción del Dr. D. Miguel Ángel Sotelo Vázquez y del Dr. D. David Fernández Llorca, la tesis doctoral titulada "**Road scene interpretation for autonomous navigation fusing stereo vision and digital maps**", cumpliéndose todos los requisitos para la tramitación que conduce a su posterior lectura.

Alcalá de Henares, June 30, 2016.

EL COORDINADOR DEL PROGRAMA DE DOCTORADO

Fdo: Dr. D. Sancho Salcedo Sanz.

Universidad de Alcalá

PhD. Program in Information and Communications Technologies

Road Scene Interpretation For Autonomous Navigation Fusing Stereo Vision and Digital Maps

PhD. Thesis Presented by

Carlos Fernández López

2016

Universidad de Alcalá

PhD. Program in Information and Communications Technologies

Road Scene Interpretation For Autonomous Navigation Fusing Stereo Vision and Digital Maps

PhD. Thesis Presented by

Carlos Fernández López

Advisors

Dr. Miguel Ángel Sotelo Vázquez

Dr. David Fernández Llorca

Alcalá de Henares, 23rd of September, 2016

A mi abuelo

"El aprendizaje es un tesoro que seguirá a su dueño a todas partes."
Proverbio chino

Quería dedicar esta tesis a mi abuelo Paco. Porque es una persona muy importante en mi vida y un referente.

Siempre le ha gustado la ciencia y de pequeño me regalaba unos libritos que venían con una revista en el que cada tomo era sobre un tema: física, anatomía, geología, tecnología, etc. Ya en aquellos años despertarse una curiosidad que me ha llevado a hacer una tesis doctoral.

Como anécdota me contaba que cuando trabajaba, estuvo ayudando a un ingeniero industrial a hacer su tesis doctoral. Mucha gente piensa que el ser doctor es cosa de médicos, pero pese a su edad, él sabia lo que era. Él contaba con orgullo su ayuda a aquel ingeniero, pero yo cuento con orgullo lo que me ha ayudado mi abuelo en la vida.

Por todos esos momentos y porque le admiro, esta tesis va dedicada a él.

Agradecimientos

Disfruta la vida,
es más tarde de lo que crees.

Proverbio chino

Parecía que este día no iba a llegar nunca, pero aquí esta. Un momento tan importante, y aquí estoy, el mismo día de enviarlo a la imprenta escribiendo los agradecimientos. Estas lineas son las mas importantes de la tesis, 'y lo sabes'. Al final es lo que todo el mundo va a leer porque del resto del libro lo que más llama la atención son las gráficas y las imágenes con muchos colores.

Quería agradecer a mis tutores Dr. Miguel Ángel Sotelo Vázquez y Dr. David Fernández Llorca todo el apoyo recibido durante estos 8 años que llevo trabajando con ellos, especialmente en los momentos mas duros de la tesis. Yo siempre veía los problemas más grandes de lo que en realidad eran y ellos me han hecho ver que al final la tesis ha quedado bastante "pintona". Gracias por ser como sois. Ese trato tan cercano hace que sea muy fácil trabajar con vosotros.

Gracias a la familia ISISLAB, porque más que un grupo de investigación somos una familia. Llevamos muchos años juntos y el ambiente de trabajo no podía ser mejor. Quería hacer un especial agradecimiento al Sr. Quintero, por sus consejos y revisiones de la tesis y por sus ideas cuando estaba desarrollando. Mucho ánimo en estos últimos meses de tesis. No te falta nada para cambiar el Sr. Por el Dr.

No me podía olvidar de nuestros vecinos de enfrente. El grupo Robesafe, que también fue mi grupo durante la primera mitad de mi andadura investigadora. Junto con el grupo ISISLAB, formamos una piña.

Muchas gracias a todos por esos momentos que hemos pasado juntos. Las conversaciones durante la comida, esos paseos que damos después de comer y por esas excursiones montañeras y gastronómicas que ya empiezan a ser tradiciones más arraigadas que la lotería de navidad. Los karts y las visitas al Jade también van por el buen camino. Hemos compartido muy buenos momentos juntos y aunque cada vez haya más candidatos a salir en 'investigadores por el mundo' sé que siempre estarán cuando se les necesite.

Gracias a los componentes que ya no están en el grupo y que se les echa mucho de menos. Fer, las comidas ya no son lo mismo sin tus temas de conversación. Óscar, echo de menos tus 'tuuu monoo', 'guachinei guachipei' y demás frases célebres. Noe, te echamos todos mucho de menos, y a tus tartas también :P. Hay muchas más miembros que dejaron la universidad, pero siguen presentes en muchas de las conversaciones y anécdotas que surgen en el día a día. (Pero en plan bien ehh, no de cotilleos y criticar).

Gracias a la CIA por no cerrarnos el chiringuito. Todos sabemos que vigilan nuestros pasos desde que le pusimos ese nombre al grupo y nuestro jefe nos llama terroristas. Algún día llegarán unos hombres de negro al laboratorio...

Gracias a mi novia Laura por aguantarme todos estos años, especialmente en el infierno que han sido estos últimos meses. Siempre había escuchado hablar de lo malos que son esos últimos meses, parecían leyendas como la chica de la curva, pero doy fe de que es verdad, yo la ví :). Laura me decía que iba muy bien, que tenía artículos publicados y estaba avanzando bastante, pero parece que la tesis no se acaba nunca. Al final tienes que poner una fecha límite porque la investigación nunca termina y una tesis nunca está acabada.

Muchas gracias a mi familia. Gracias por todo el apoyo recibido durante estos años de investigación. Gracias por esas preguntas de ¿qué tal va la tesis?. Porque aunque no os estuviérais enterando de lo que os estaba contando, me ayudaba mucho a desahogarme y eso ha sido muy importante.

Gracias a los amigos de toda la vida Álvaro y Cristina. Gracias a la familia del roco por esas tardes de risas que hemos pasado mientras escalábamos. Me han ayudado mucho a desconectar del estrés de la tesis.

Gracias a esta tesis, porque sin ella no habría tenido la posibilidad de haber viajado por todo el mundo. La excusa de los congresos me ha dado la oportunidad de conocer China, Korea y Canarias. Cuando obtenía unos buenos resultados siempre preguntaba con una gran sonrisa ¿dónde es el siguiente congreso?.

Parece que esto se acaba. La tesis es uno de los hitos en la vida que ya he cumplido. Ahora me queda una sensación agridulce y la pregunta: ¿y ahora? ¿Qué voy a hacer sin la tesis? Pues vivir hombre! Disfrutar de las pequeñas cosas de la vida, de los amigos y la familia (especialmente de mi sobri), que en estos últimos meses no he podido disfrutar de ninguna de ellas.

Resumen

En esta tesis se ha presentado un método de detección de carretera basado en visión estereoscópica. El aprendizaje automático se utiliza para resolver problemas de visión artificial de muy diferente ámbito, en concreto, la técnica utilizada en este caso es la llamada *boosting*, la cual utiliza árboles de decisión para clasificar cada píxel de la imagen como zona que pertenece carretera o no. El vector de características utilizado incluye información proporcionada por mapas digitales, visión estéreo y cámaras en color y en escala de grises.

La imagen en escala de grises es utilizada para detectar marcas viales, Local Binary Patterns (LBP) y Histogramas de Orientación de Gradiente (HOG). Las cámaras en color son utilizadas para el cálculo de una imagen que es invariante a la iluminación y también para detectar las sombras presentes en la imagen. Además, se ha desarrollado un método basado en el espacio de color HSV para detectar las zonas de vegetación presentes en la escena. Las cámaras estéreo tienen un papel importante porque son las encargadas de proporcionar información 3D al sistema. Algunas de las características que usan dicha información son los vectores normales y los valores de curvatura.

Se ha desarrollado un nuevo método para la detección de bordillos. Este novedoso detector de bordillos se basa en el análisis de la curvatura porque describe la variación de la forma de la carretera incluso en presencia de pequeños bordillos. La función es capaz de detectar bordillos de 3 cm de altura incluso hasta 20 metros de distancia, siempre y

cuando los píxeles que pertenecen al bordillo estén conectados entre si en la imagen de curvatura. Otros obstáculos como vehículos, muros o arboles son también detectados utilizando visión estereoscópica.

Una nueva forma para convertir características que describen limites de carretera en características que describen zonas de carretera se ha descrito en esta tesis. Utiliza marcas viales, bordillos, obstáculos y zonas de vegetación como entradas y tras incluir información adicional del mapa se genera un modelo de carretera. La originalidad de este sistema es el punto desde donde se detecta es espacio libre.

Otra característica muy importante es la obtenida a partir de los mapas digitales. El objetivo es conseguir un imagen a priori de la forma de la carretera basado en la posición actual del vehículo y la información de las calles proporcionada por el mapa. La incertidumbre sobre los errores de posicionamiento son tenidos en cuenta durante el proceso y la anchura de la carretera es correctamente detectada usando el modelo radial propuesto.

Se han realizado múltiples pruebas con diferentes clasificadores y parámetros basados en arboles de decisión para posteriormente elegir el clasificador que mejor funciona en la detección de carretera. El resultado de la clasificación es utilizado en un CRF para filtrar la respuesta y obtener un resultado mas suave.

La métrica utilizada para evaluar los clasificadores es el $F_1 - score$. El sistema es evaluado en el plano imagen, el cual es el método mas común en la literatura. Sin embargo, en un escenario de conducción autónoma, el control se realiza normalmente en una imagen a vista de pájaro de la escena. Se ha adoptado el mismo método de evaluación que se utiliza en la comparador internacional de algoritmos KITTI para poder comparar nuestros resultados con otros algoritmos.

Palabras clave: espacio libre, detección de carretera, bordillos, vehículos autónomos.

Abstract

A stereo vision based road detection method is presented in this thesis. Machine learning is widely applied to solve computer vision problems, particularly, the applied technique is boosting, which internally uses decision trees to classify every pixel of the image as road or non road pixels. The feature set includes information provided by a digital navigation map, 3D stereo vision, color and grayscale cameras.

The features obtained from the grayscale camera are road markings, Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). Color cameras provide an illuminant invariant image and a shadow detection function. In addition, Hue Saturation Value (HSV) and a vegetation detector is developed to discard green areas of the scene. The stereo cameras have a very important role because they supply 3D information to the system. Some features that take advantage of 3D points are normal vectors and curvature values.

A novel method for the specific task of curb detection is developed. The basis of the curb detector is the curvature feature because it describes the variation of the road shape even in presence of small curbs. The function is able to detect curbs of 3 cm height up to 20 meters whenever the curb is connected in the curvature image. Other obstacles such as vehicles, walls or trees are also detected using stereo vision methods.

A new approach is created to merge features that describe road limits to a new feature that describes road surface. The new feature uses

road markings, curbs, obstacles and vegetation areas to obtain a road model with the additional information of the number of lanes provided by the digital navigation map. The originality of the presented method is the point where the road limits are detected from. Other methods create radial rays from the bottom center of the image until the ray reach an obstacle. Our proposal finds the road limits from a different point of view, its rays start from the vanishing point along the image and their accumulated values are analyzed to obtain a road model.

Another important feature is obtained from a digital navigation map. The aim is to get a prior model of the road based on the GPS position and the information extracted from a digital map. The uncertainty around the correct position is taken into account during the modeling process and the road width is precisely adapted thanks to the radial ray road model.

Several tests have been deployed with different classifiers based on decision trees to choose the best type of classifier. After all, the features mentioned before are integrated into a boosting classifier. It generates a probability to be road for every pixel, which is used into a Conditional Random Field (CRF) to filter the classifier response and obtain a smoother result.

The metric for the classifiers evaluation is the $F_1 - score$, which is the harmonic mean of precision and recall. The system is evaluated in the image plane, which is the most common approach in the literature. However, in a vehicle scenario, its control stage usually happens in a 2D Bird's Eye View (BEV). The KITTI benchmark has a ranking sorted by $F_1 - score$ calculated on the BEV images. In order to compare our system with other algorithms in an international benchmark, the same evaluation method is adopted.

Keywords: free space, road detection, curb, autonomous vehicles.

Contents

List of Figures

List of Tables

List of Acronyms

ADAS	Advanced Driver Assistance System.
BEV	Bird Eye View.
CNN	Convolutional Neural Network.
CRF	Conditional Random Fields.
DARPA	Defense Advanced Research Projects Agency.
DEM	Digital Elevation Map.
DT	Decision Trees.
EM	Expectation Maximization.
ERT	Extremely Randomized Trees.
EU	European Union.
FP	False Positive.
GB	Gigabyte.
GMM	Gaussian Mixture Model.
GPS	Global Positioning System.
GPU	Graphical Processing Unit.

HOG	Histograms of Oriented Gradients.
HSV	Hue Saturation Value.
IMU	Inertial Measurement Unit.
IoT	Internet of Things.
LBP	Local Binary Pattern.
LIDAR	Light Detection And Ranging.
LMEDS	Least Median of Squares.
M-SAC	M-estimator SAmple and Consensus.
ML	Machine Learning.
MRF	Markov Random Fields.
NHTSA	US National Highway Traffic Safety Administration.
NN	Neural Networks.
OSM	Open Street Maps.
PCA	Principal Components Analysis.
RADAR	Radio Detection And Ranging.
RANSAC	Random Sample Consensus.
RMS	Root Mean Square.
RT	Random Trees.
SGM	Semi Global block Matching.

SVM Support Vector Machine.

TP True Positive.

UM Urban Marked road.
UMM Urban road with Multiple Marked lanes.
US United States of America.
UU Urban Unmarked road.

V2I Vehicle to Infrastructure.
V2V Vehicle to Vehicle.

Chapter 1

Introduction

1.1. Context Analysis

Transportation has an important role in the economy and provide benefits to the society in terms of quality of life. However, every year worldwide organizations and governments publish studies about road accidents and their consequences. This is the case of [1]. It affirms that traffic accidents kill 1.25 million people a year worldwide. Despite this dramatic statistic, human is currently the absolute best driving machine, even when human error still accounts for 90% of all road accidents. The high number of casualties on the road can be explained by many factors. As reported in [2], more than 12.000 lives can be saved per year on european roads if everybody fasten their seat belt, respect speed limits and do not drive under the influence of alcohol.

Distraction is another factor since drivers need to keep their attention focused on surrounding traffic continuously, not just for their own safety but for the sake of their passengers and other road users too. Visual distractions make drivers take their eyes off the road. Cognitive distractions cause drivers to think about other things and manual distractions cause drivers take their hands off the wheel. Often, many distractions happen at the same time. For example, a driver that has

to turn round to deal with kids fighting in the back will not have their eyes on the road or their attention focused on the traffic. Mobile phones and satnavs are major sources of distraction. The use of hands-free and hand-held phones produce similar impairment in performance compared to normal driving without using a phone [3]. The driver's response to critical events is impaired more than the ability to maintain vehicular control.

Apart from driver distractions, every road elements (vehicles, drivers, infrastructure) play an important role in the probability of crash or the final outcome. For this reason, new technologies are also applied to vehicles and infrastructure. Connected vehicle research is a multimodal initiative that aims to enable safe, interoperable networked wireless communications among vehicles, the infrastructure and passengers' personal communications devices. The vision for connected vehicle technologies is to transform surface transportation systems to create a future where:

- Highway crashes and their tragic consequences are significantly reduced.
- Traffic managers have data to accurately assess transportation system performance and actively manage the system in real time, for optimal performance.
- Travelers have continual access to accurate travel time information about mode choice and route options, and the potential environmental impacts of their choices.
- Vehicles can talk to traffic signs to eliminate unnecessary stops and help drivers to operate vehicles for optimal fuel-efficiency.

Governments are concerned not only about traffic accidents. Moreover, optimal fuel efficiency is one of the most important objectives for the governments because greenhouse gas emissions are one of the main causes of climate change and road transport contributes about

one-fifth of the EU's total emissions of carbon dioxide (CO_2), the main greenhouse gas [4]. While these emissions fell by 3.3% in 2012, they are still 20.5% higher than in 1990. In fact, transport is the only major sector in the EU where greenhouse gas emissions are still rising, see Figure 1.1. The target for 2015 requires that new cars do not emit more than an average of 130 grams of CO_2 per kilometer.

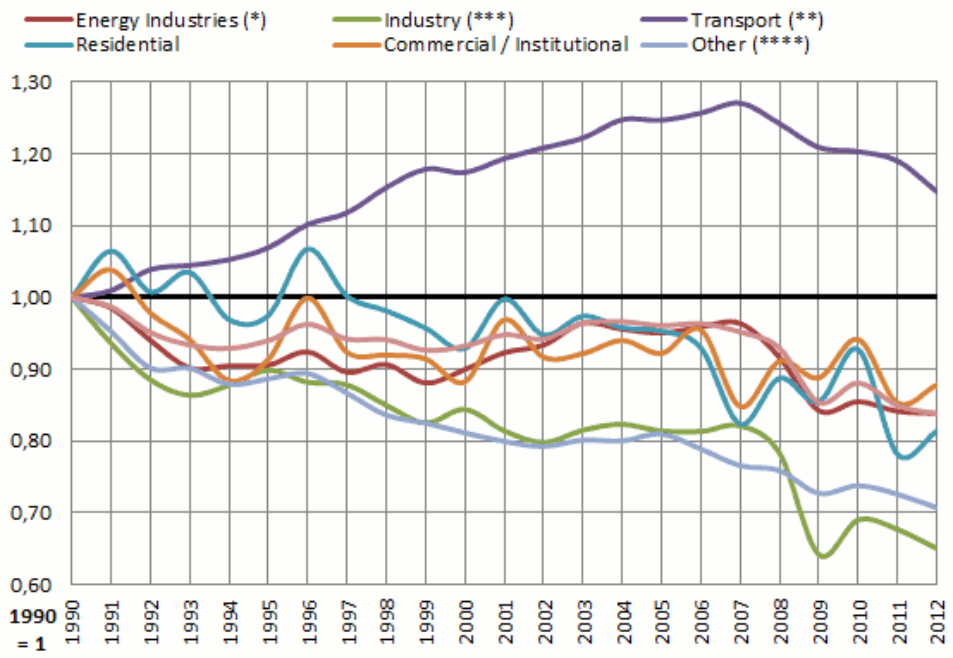


Figure 1.1: EU greenhouse gas emissions from transport and other sectors, 1990-2012. (*) Excluding LULUCF (Land Use, Land - Use Change and Forestry) emissions and International Bunkers (**) Excluding International Bunkers (international traffic departing from the EU) (***) Emissions from Manufacturing and Construction and Industrial Processes (****) Emissions from Fuel Combustion in Agriculture/Forestry/Fisheries, Other (Not elsewhere specified), Fugitive Emissions from Fuels, Solvent and Other Product Use, Waste, Other.

Future scenarios aim to increase the efficiency in several aspects. Autonomous driving can help to reduce the number of accidents and also the CO_2 emissions. The automation of vehicles is growing year by year. According to a U.S. Department of Transportation report [5], 55% of US road accidents are caused by a vehicle leaving its lane, thus, lane keeping assistant systems can help to prevent this. As automa-

tion technology advances, all of these distractions and human wrong decisions can be eliminated by means of automation technology advances. Thereby, automated driving can save millions of lives, in fact, the adoption of automated cars could reduce accidents by 80% by the year 2040.

In terms of CO_2 emissions, a Nature climate change study [6] estimates that emissions per mile from light-duty vehicles could fall by as much as 94% by 2030 in a best-case scenario of electric driverless taxis. However, an automated transportation might increase fuel use because driverless technology might include people who do not usually drive as potential users. Some benefits for this new users are personal independence, reduction of social isolation and access to essential services [7]. In addition, autonomous vehicles can increase fuel efficiency due to Vehicle to Vehicle V2V and Vehicle to Infrastructure V2I communications. Thanks to communications, cars can flow right by each other at intersections, avoiding stopping and starting. An experiment of these communication techniques on traffic lights has demonstrated a reduction of 15% of CO_2 emissions in urban traffic [8]. Driving in highways can take advantage of platooning, which allows vehicles to drive directly behind another one by means of a communication protocol, reducing the air drag and saving fuel and CO_2 emissions.

The productivity can be increased with autonomous vehicles. According to a study [9], Americans waste 111 hours annually per driver. With the introduction of automated vehicles into daily lives, the passenger suddenly becomes far more appealing to a whole new world of business opportunities happening right inside the car. The passenger will spend time web surfing, reading, enjoying multimedia entertainment, accessing social media, sending/reading emails, doing office work, taking a nap, etc. As a consequence, the productivity will be significantly boosted.

The number of vehicles is rising and it is required the creation of smart routing systems that can make moving around more efficient, re-

ducing congestion and increasing the potential number of vehicles per kilometer on roads. Currently, most vehicles spend 90-95% of their time not operating. As a consequence, fast growing shared mobility initiatives are creating diverse dynamics worldwide by offering multiple options. If a vehicle can be used 60-70% of the time through autonomous vehicles being flexible in their characteristics, less people will need to own a vehicle [10].

Another interesting future technology is the Internet of Things (IoT), which is getting popular and the number of devices connected to internet is increasing day by day. The vehicles of the future will be another device that receives information from many sources: the infrastructure, other vehicles, pedestrians, traffic congestions, bicycles, etc. The human driver is not able to manage this amount of information properly. The decision making task becomes much more efficient when the driver is an autonomous vehicle.

1.2. Motivation

Autonomous vehicles require a precise and robust perception of the environment. It is a crucial point in the development of autonomous vehicles because the perception layer is the base for higher level systems, such as control algorithms or path planning. One of the main issues is the road detection. It has traditionally been an exhaustive topic of research in the fields of Advanced Driver Assistance System (ADAS) and autonomous driving. On the one hand, ADAS have mainly focused on increasing the safety of drivers and road users by means of warnings to drivers and assisted interventions. On the other hand, it is undebatable that autonomous driving has become a high priority issue on the research and commercial agendas of major car makers in the latest years, aiming at producing fully autonomous vehicles by 2020. The deployment of autonomous cars will bring a number of clear benefits in terms of increased traffic efficiency and reduced accident toll,

deriving in unquestionable higher energy efficiency and enhanced road safety.

The Grand Challenge organized by DARPA in 2005 was the first championship of autonomous vehicles. The participants relied on precise and expensive differential GPS and IMU sensors to follow a set of waypoints along a planned route. Dynamic obstacles detection was solved thanks to a multi beam LIDAR. This approach has demonstrated its ability to drive safely in highways and urban environments [11–13], however a high definition map is required. In Figure 1.2, the enriched map integrates precise information of lane markings, curbs, intersections, buildings, etc. The drawback of this type of maps is their size ($\sim 2GB/km$), the complex process to integrate all the measurements and how to update the map. The price of the LIDAR ($\sim 75K$) and the GPS+IMU($\sim 25K$) is not affordable for the car industry.

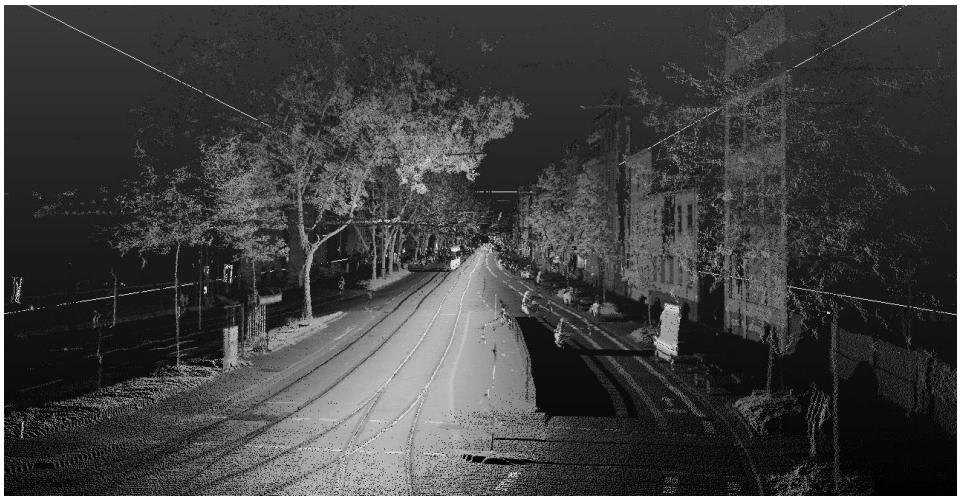


Figure 1.2: High definition scene reconstruction after the integration of several measurements of a multi beam LIDAR.

For this reason, our proposal of road detection is based on a low cost GPS sensor and a pair of stereo cameras ($\sim 2K$). The goal is the creation of a method that uses relative low cost sensors and detect the road in the most challenging scenarios. Our system aims to be part of a local navigation system with a global route planner. It tries

to imitate the human way of driving: a user wants to drive from A to B and a global route planner uses a low cost GPS sensor to locate the vehicle and plan the route. When the route is ready, local navigation is required to detect the drivable area and take the corresponding decisions.

1.3. Challenges

A computer vision approach for road detection presents several challenges due to the large variability of road appearance arising from the different road types. Furthermore, lighting variation and weather conditions create shadows and reflections, which are one of the most difficult scenarios for computer vision algorithms.

In Figure 1.3, a collection of some challenging scenarios is presented. In Figure 1.3a, a residential area has parking spots with the same texture of the road and a sidewalk with a different one. The only difference is a small curb with another texture. Furthermore, the road has two different textures: asphalt and bricks. Contrary to Figure 1.3a, in Figure 1.3b the sidewalk and the road have the same color, texture and nearly the same height. This scene claims to use high level context information, it means that the road should be the space between the rows of bricks with different texture. Figure 1.3c shows how scenes with high illumination contrast usually saturate some pixels and makes more difficult the correct segmentation of the road. In addition, the shadows of buildings and trees create many edges, requiring too much effort to mitigate them. Finally, in Figure 1.3d another scene is affected by shadows. Darkness produces that the road and the sidewalk on the left have pixels with similar values.

In addition, the use of stereo vision provides 3D information of the scene. This feature is very useful to detect obstacles and complement the texture analysis. However, the disparity map is affected by mismatching errors and 3D points are not always reliable.

(a) Challenging scenario with different textured roads.

(b) Challenging scenario where road and sidewalk have the same texture and nearly the same height.

(c) Challenging scenario with pixels close to the camera saturation and irregular shadows due to trees.

(d) Challenging scenario with very dark pixels. The curb on the left is difficult to distinguish from the road.

Figure 1.3: Collection of some challenging scenarios in urban environments.

1.4. Applications

A correct drivable space detection is the base for many applications. The knowledge of the scene is the most important task for a correct decision making system. The first and the most obvious application is autonomous vehicles navigation. From this point other applications rise. For example, if the road is detected and a pedestrian is crossing the street, the autonomous vehicle can brake to prevent an accident. If the pedestrian is close to the vehicle but on the sidewalk, the vehicle could drive safely along the planned route.

Another application of road detection is map updating using floating vehicles. This application is very useful to maintain collaborative navigation maps up to date thanks to several vehicles feedback.

As part of the road detection, road markings can be identified and a score of how good those are painted can be sent to the maintenance center and repaint areas where the quality is deficient.

Finally, if the free space is precisely recognized, free parking spots can be detected. The position of the spot could be sent to a control center and other users can find a place to park close to their destination.

1.5. Document Outline

After the presented introduction, the remainder of the document is organized as follows. Chapter 2 contains a brief review of the published research on road detection algorithms. In particular about sensing and geometrical modeling of the road. Afterwards, different features integration methods are reviewed. Chapter 3 describes the developed road detection method, based on feature analysis and machine learning classifiers. In chapter 4, the results of the algorithm are presented and discussed. Chapter 5 contains the conclusions and main contributions of the thesis. Finally, chapter 6 contains future research lines that may spring from it.

Chapter 2

State of the Art

As mentioned in section 1, autonomous vehicles are going to change the mobility around the world. This technology is divided in 4 or 5 different levels of automation depending on the organization. The Germany Federal Highway Research Institute (BASt) and the US National Highway Traffic Safety Administration (NHTSA) has established 4 levels of automation, however SAE International summarizes 5 levels of automation:

- Level 0: No automation. The full-time performance by the human driver of all aspects of the dynamic driving task. No systems intervene - only those that warn the driver. The driver must constantly monitor the drive.
- Level 1: Driver assistance. The driving mode-specific execution by a driver assistance system of *either* steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the dynamic driving task. The driver must constantly monitor the drive. He must be ready to resume full control immediately.
- Level 2: Partial Automation. The driving mode-specific execu-

tion by one or more driver assistance systems of *both* steering *and* acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task. The driver must constantly monitor the driver. He must be ready to resume control immediately.

- Level 3: Conditional Automation. The system takes over both steering and acceleration / deceleration in a defined use case. It is capable of recognizing its limits and notifying the driver. The driver does not need to monitor the drive, but be ready to resume control within a given time frame if the system so requests.

- Level 4: High Automation. The driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene. The driver can hand over the entire driving task to the system in a defined use case. The driver would not be required at all during these cases - neither for monitoring, nor as backup.

- Level 5: Full Automation. The full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver. The driver is no longer required at all.

How can we reach level 5 of autonomous driving? The three pillars of autonomous driving are sensing, mapping and driving policy (path planning). The sensing interprets the scene with 360° awareness and produce an environmental model. The environmental model includes where the moving obstacles are, road limits, curbs, barriers, etc. The sensing is based on cameras, RADAR and LIDAR sensors. Cameras provide rich information of the scene with high frequency, however they are affected by illumination and weather conditions. RADARs

are more robust to dust and other weather conditions, however they cannot sense texture.

Mapping, either as a part of sensing or a layer redundant to sensing, requires some sort of connectivity for the purpose of updates. There are two main types of maps, the first type are navigation maps, which provides information about the steps to reach our destination. The second type are high definition maps, which provides 3D information of the environment with centimeter precision.

The last pillar is driving policy, which includes the set of rules to merge in the traffic and manage driving behaviors. The driving policy needs to learn human driving behaviors in order to drive properly in mixed traffic of human driving and autonomous driving.

This chapter is organized as follows: In section 2.1, a collection of sensors applied to environment perception are presented and their principal applications are analyzed. The resulted detection can be modeled using different geometrical models, which are explained in section 2.2. Section 2.3 presents different machine learning techniques used for road features integration. Conclusions related with the analysis of the state of the art are presented in section 2.4 and the chapter concludes with the main contributions of this thesis in section 2.5.

2.1. Sensing

The sensing interprets the scene with 360° awareness and produce an environmental model. It can be achieved using different types of sensors. These can be active and passive sensors. On the one hand, active sensors work at long distances and under bad weather conditions. Some examples are LIDAR and RADAR, which are able to detect obstacles at more than 100 meters. On the other hand, passive sensors have the main advantage of their low cost. Furthermore, visual information can be very important in some applications such as traffic

sign recognition or object identification. Figure 2.1 shows a general classification of different approaches for road detection depending on the sensor and the methodology.

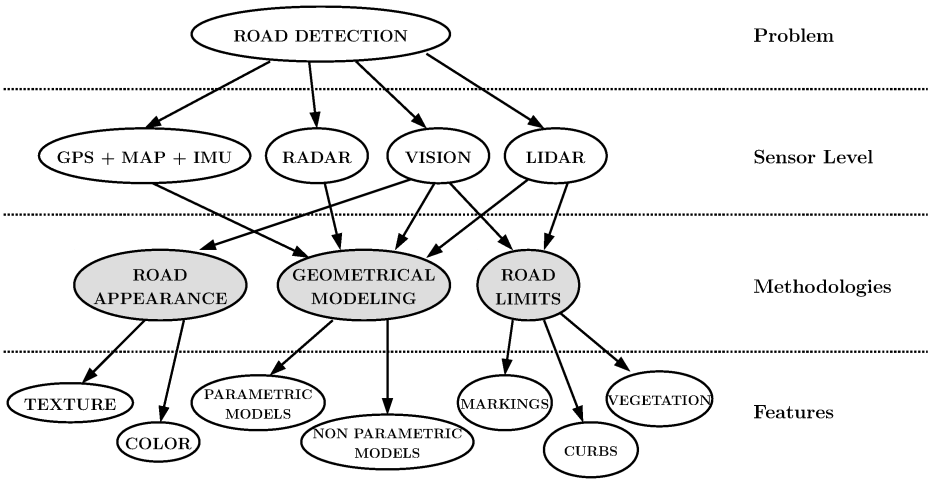


Figure 2.1: Classification of different road detection approaches depending on the sensor and the methodology.

LIDAR technology has been widely used to detect curbs, the road surface and different types of obstacles. These can be classified into two types: single beam (2D) and multi beam (3D). 2D LIDARs are usually mounted on the front of the vehicle parallel to the ground plane for obstacle detection. This approach is affected by the pitch variations caused by the vehicle movement. It provokes the system to detect the road surface as an obstacle when the sensor is installed close to the road. On contrary, if the sensor is installed in a higher position, obstacles below its height are not detected, see Figure 2.2.

When the goal is road surface or curb detection, the sensor is mounted on the top of the vehicle facing down. This configuration is good to detect height variation on the road shape, however, the shape is only detected in a fixed distance. For this reason, it is important to integrate the measurements along the time compensating the ego-motion of the vehicle. Thanks to measurements integration,

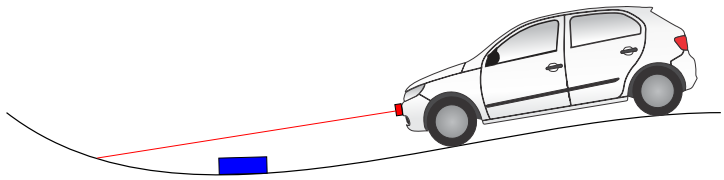


Figure 2.2: 2D LIDAR installed in a vehicle front bumper. The LIDAR beam (in red) passes over the obstacle (in blue). In addition the road is detected as an obstacle when the vehicle is driving on a non flat surface.

the confidence on the scene interpretation is increased and road shape can be modeled using clothoids, splines or other type of model. This configuration is also valid for road marking detection because LIDARs provide distance information and also surface reflectivity intensity.

Beside, there are LIDARs with 4, 6, 16, 32 and 64 layers. The resulted point cloud is sparse compared with a single beam LIDAR with temporal integration, however, they create a new scenario of possibilities because they provide 360° precise 3D information of the environment from a vertical field of view of 26°, generating a point cloud of the entire scene, see Figure 2.3. They have been used for many tasks, including road marking [14, 15], road [16, 17], curb [18–22] and vehicle detection [23–25].

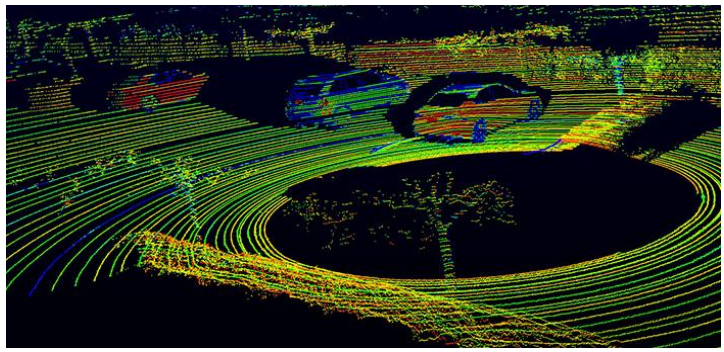


Figure 2.3: Single spin of a multi beam LIDAR. The precision of the acquired information and the vertical resolution make scene interpretation more robust than single beam approaches.

Different methodologies to detect the road can be distinguished. Some of them are based on road appearance learning, where the main

features are texture and color information. The second approach is focus on road limits detection, assuming that the space between limits is the road surface. Finally, modeling tries to extract a compact high level representation of the road. In order to clarify the classification, Table 2.1 summarizes different methodologies of road detection.

Methodology	Sensor	Feature	Examples	Refs
Road Appearance	Vision	Texture	HOG, anisotropy, filter bank, LBP	[26–31]
		Color	Illuminant invariant, GMM	[32–37]
Geometrical Modeling	Map+GPS RADAR Vision LIDAR	Parametric Models	Parabolic curves, clothoids, B-splines, snakes	[38–42]
		Non Parametric Models	Ant Colony Optimization, Dijkstra, A^*	[43–45]
Road Limits	Vision LIDAR	Road Markings	Reflectivity intensity, adaptive threshold	[46–51]
		Road Curbs	Digital Elevation Maps, Difference of heights, curvatures	[18–22, 52–65]

Table 2.1: Classification of different road detection methodologies.

2.1.1. Road Appearance

Road appearance is the feature that makes vision sensors really relevant. When somebody asks you to describe an object, you probably start estimating its size, then color, and finally the texture. The difference between the object texture/color and the background determines the object shape. This procedure is also applied to detect the road, however, this task is a very complex task because there are several textures and colors for the same object. Figure 1.3 shows some examples of different scenarios where the road is completely different from one situation to another.

This challenging problem is managed in the state of the art model-

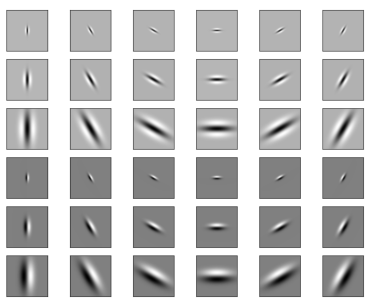
ing road pixels histogram with a Gaussian Mixture Model (GMM). The models are usually estimated using some optimization method such as Expectation Maximization (EM). This method is used in [32,33], however the shadows make the system to fail and classify shadowed areas as non road. In order to solve this problem, [34] uses a flexible number of color models in a modified Hue that is invariant to brightness [35]. Another strategy is to remove shadows from images [36] by converting the original RGB image into an illuminant invariant image. The resulting image is used in [37] to detect the homogeneous road surface.

Texture analysis is used to obtain a descriptor of the surface. On the one hand, Gabor filters [26,27] and Histograms of Oriented Gradients (HOG) [28] provide information about the orientation. However, strength of texture anisotropy describes the homogeneity in a region of interest [29,30]. More information about road appearance is extracted from a filter bank, which is an array of band-pass filters that usually have different scales and orientations. In Figure 2.4, a collection of some road appearance features is displayed to show their behavior on different situations. Finally, another common descriptor is Local Binary Patterns (LBP) [31], which describes the relation between the evaluated point and its surrounding values.

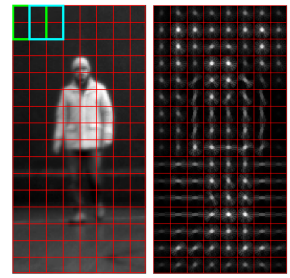
Those features per se are not robust for a road detection method, however they are usually included into a larger feature vector for a machine learning strategy. Some machine learning techniques will be explained in section 2.3. Moreover, the computer vision features can be combined with other sensors such as LIDAR to detect the road and achieve improved results [66, 67].

2.1.2. Road Limits

This section presents an opposite approach comparing with the road appearance approach. The goal is to detect the features that describe the road limits. In most of the cases the road is limited by curbs,

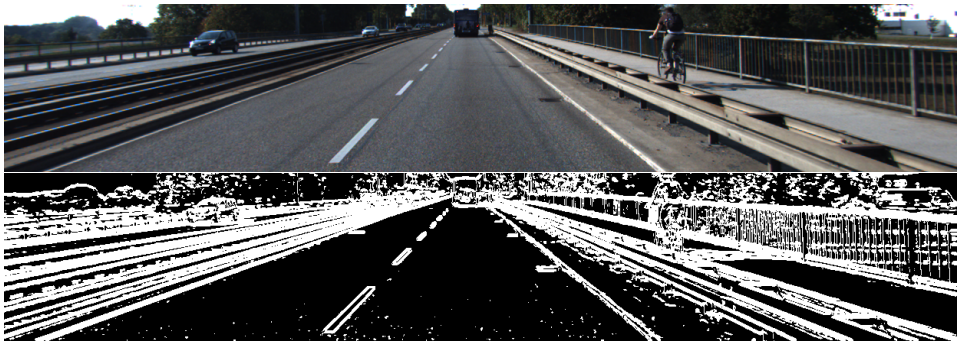


(a) Filter bank.



(b) HOG applied to pedestrian detection.

(c) Illuminant invariant image.



(d) Strength of texture anisotropy.

Figure 2.4: Different road appearance descriptors.

road markings, vegetation areas or parked cars. These features can be estimated using computer vision, RADAR, LIDAR or a combination of all of them.

RADAR sensors have been used to detect vehicles in the last decades [68,69]. Actually, RADAR sensors are consolidated in common vehicles for Automatic Cruise Control (ACC). Driving in highways requires a

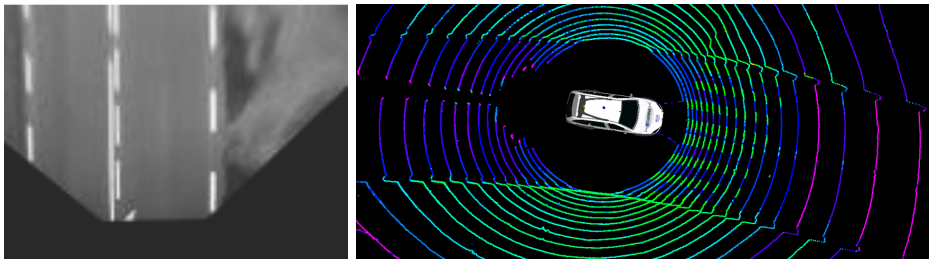
long detection range due to the vehicles high speed. RADAR sensors are adequate for this task since their range reach 150-200 meters. They are also applied to detect the road [70–73], however their use for that is not as extended as their use for vehicle detection. Some approaches fuse RADAR jointly with vision sensors to improve detection and tracking results for vehicles [74–77] and road [78, 79].

2.1.2.1. Road Markings

Current ADAS have integrated road marking detection in highways because it is the most important feature to keep the vehicle into the current lane. Due to the contrast between the road and the lane marking, the problem is addressed by looking for gradients in the image. Three different methods are usually deployed. The first one consists of a median filter applied to the original image to remove some noise. Afterwards, a filter is applied to boost transitions of dark-light-dark pixels [46]. The further lines are thinner than the closer ones due to perspective. For this reason the filter size should be adapted with the distance. This drawback can be avoided making a perspective transform as shown in Figure 2.5a. Thanks to the constant scale along the distance in the resulting image, the filter keeps the same size.

The second method consists of a median filter twice the size of the line size. The resulting image is a smooth image without lane markings, which is subtracted to the original image. The result is an image with road markings highlighted [47]. The negative point of this method is that zebra crossings or other lane markings wider than the line are partially detected. The third method consists of an adaptive threshold to separate the dark pixels of the road from the light ones of road markings [48].

Road markings are painted with reflective materials to make them highly visible during the night. That property makes them detectable with a LIDAR sensor because they detect surface reflectivity intensity



(a) Bird Eye View (BEV) of a road extracted from [80].

(b) Zebra crossing detection (blue and green stripes) with a 3D LIDAR extracted from [49].

Figure 2.5: Road marking detection using different sensors.

[49–51]. In Figure 2.5b, an example of road marking detection is shown. If a 2D LIDAR is used for this task, a temporal integration is required to increased the number of measurements and make the detection more robust. However, using 3D LIDARs the integration is not compulsory because the resulting point cloud has a sufficient number of points to detect the lines.

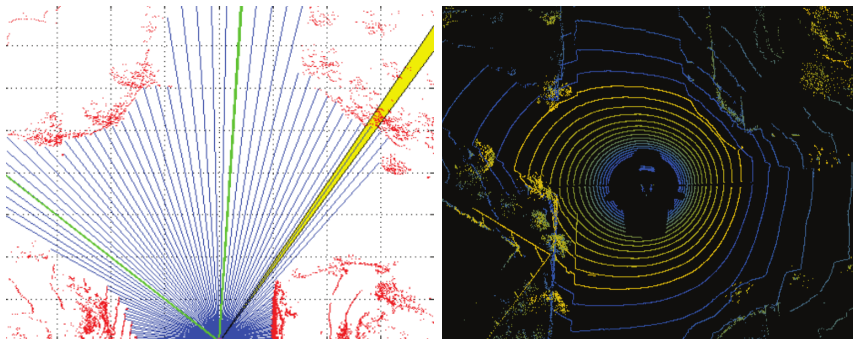
In addition to road marking detection, the correct interpretation of arrows and other symbols can be taken into account in higher level modules for a correct navigation [81, 82].

2.1.2.2. Curbs

Road markings are present is most of the roads, however in rural roads and residential areas might be not present. In these scenarios, road curbs are also a discriminant descriptor to delimit the drivable area, specially in urban environments. Depending on the city, curbs have very different sizes: from 3 cm up to 12 cm or even more. Specific algorithms have been developed to solve this challenging problem. There are two main approaches depending on the sensor. The first one is using LIDAR and the second one is using cameras. Stereo cameras depend on image pair matching methods to obtain depth information. Although 2D LIDARs are able to directly return this information, only few curb points can be detected using this sensor [52–54]. 3D LIDAR

provides a dense point cloud and thus makes possible to detect a larger extent of the curb [18, 22, 55, 56].

LIDARs provide precise measurements, which is a very important feature to detect small curbs. The most common features are elevation difference, gradient value and normal orientation. They can be estimated analyzing the measurements as rings or as longitudinal rays, see Figure 2.6.



(a) Measurements analyzed as longitudinal rays extracted from [56].

(b) Measurements analyzed as rings extracted from [22].

Figure 2.6: Curb detection using a 3D LIDAR.

Detection distance using LIDAR is up to 50 meters [21]. On the contrary, it is not possible to reach that rate with a pair of stereo cameras, which have a detection range of 20 meters approximately. The most common approach to detect curbs in the literature is the use of a Digital Elevation Map (DEM) to integrate the 3D measurements and a posterior analysis of the height variation [57–61]. Some of them try to model the curb shape with cubic polynomials [60] or a cubic spline [61] however the diverse type of shapes present in urban scenarios make those methods fail in some scenarios. A combined detection using computer vision and LIDAR sensors is presented in [62–64] with accurate results thanks to the complementary features of each sensor.

The principal parameters for the algorithm evaluation are the curb height and lateral distance with respect to the ground truth. On the one side, in [57, 59] the minimum curb height detected is 5 cm with

an error of 3 cm at 20 meters. On the other side, in [58] the lateral distance is evaluated up to 20 meters with an error of 20 cm.

Our novel method based on curvature values is presented in [65], where the system is evaluated on point clouds from stereovision and 3D LIDAR. The method obtains a lateral error of 14 cm at 20 meters distance and is able to detect curbs of 3 cm height up to 20 meters whenever the curb is connected in the curvature image. This method is described in detail in section 3.2.3.3.

2.2. Geometrical Modeling

2.2.1. Parametric Models

The road and lane detection are usually guided in a top-down manner by fitting a geometric model to the visual features extracted from the image. The goal is to extract a compact high level representation of the path. The simplest geometric model used for road boundaries are straight lines. Due to the pin hole camera model, straight parallel lines converge in a vanishing point. This principle is exploited in the state of the art to detect the road using an edge descriptor extracted from the image [26, 83, 84]. The edge descriptor is usually a Canny or Sobel filter. Afterwards, straight lines are fitted using the Hough transform. The intersection point of all the lines is the vanishing point, see Figure 2.7. It can be estimated using LMEDS, RANSAC, M-SAC or other fitting method.

Roads are not always straight. For that reason, a modified version is presented in [85], where 4 different vanishing points are estimated at 4 different distances. The result is a sequence of straight lines that model the road shape. More complex models are used to model curved roads, such as parabolic curves [38], clothoids [39], B-splines [40, 41] or active contours (snakes) [42].

These parametric models improve the noisy bottom-up detections

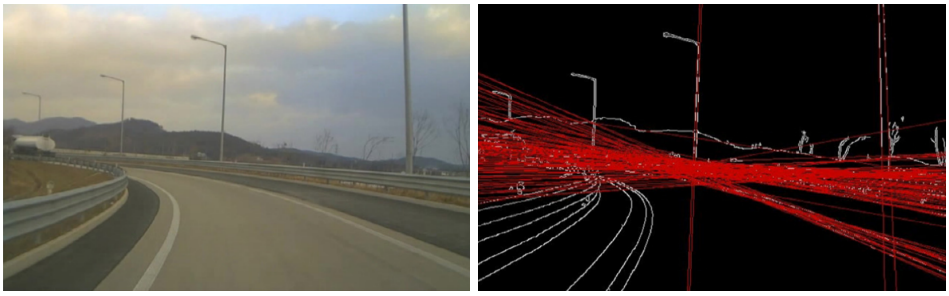


Figure 2.7: Images of the vanishing point estimation extracted from [84].

due to their constraints of width and curvature. Urban environments are more difficult to model because the presence of parked cars do not fit to some of the width restrictions.

2.2.2. Non Parametric Models

Non parametric models are less common because they demand only that the line be continuous. It provokes the model to be less robust than parametric models but more flexible to adapt to the irregular shapes present in urban environments [43] or rural paths [44]. One example of a non parametric model is Ant Colony Optimization (ACO) for finding optimal trajectories on the image plane [45]. In addition, the image can be formulated as a graph, and traditional graph algorithms such as shortest path, Dijkstra or A^* can be applied to obtain the continuous road boundary.

2.2.3. Map Based Models

High definition maps are a robust way to navigate [11–13]. They are usually built integrating several measurements of a multi beam LIDAR [86, 87] or multiple single beam LIDARs [88] and their size is $\sim 2GB/km$, which is difficult to manage in a long trip or in a city. Tom Tom, Nokia's HERE Maps and Google Maps are working on these types of maps. During 2015 HERE maps was sold to Audi, BMW and

Mercedes. High definition maps require precise localization ($\sim 10cm$), whereas navigation maps do not require that precision ($\sim 10m$).

The ultimate goal is to drive everywhere with full functionality, but there are two main points of view how to get to it. The first approach is followed by technology companies (Google, Baidu, etc.). This point of view tries to drive in some places with full functionality using 3D detailed map ($\sim GB/km$) and low resolution sensing (LIDAR). On the other side, car industry tries to drive everywhere with partial functionality using low resolution maps and high resolution sensing. The problem of the first approach is the scalability of the map and the updates. The update of the map is safety critical because it is used to navigate. The problem of the second approach is to get stronger artificial intelligence. The ultimate goal is to get cognitive perception as humans do, however without that level of artificial intelligence, a higher resolution map is required to compensate that weakness.

The higher resolution map can be updated beyond sensing. Instead of high resolution maps ($\sim GB/km$), sparse 3D maps can be built using landmarks and they can be stored in ($\sim KB/km$). These maps can be shared and the map can be updated using crowd sourcing. Toyota's newly developed system uses automated cloud-based spatial information generation technology to generate high precision road image data from the databanks and GPS devices of designated user vehicles. While a system relying on cameras and GPS in this manner has a higher probability of error than a system using three-dimensional laser scanners, positional errors can be mitigated using image matching technologies that integrate and correct road image data collected from multiple vehicles, as well as high precision trajectory estimation technologies. This restricts the system's margin error to a maximum of 5 cm on straight roads. The problem of this approach is the use of bandwidth to send the images to data centers if many vehicles include the technology. However, to support the spread of automated driving technologies, Toyota plans to include this system as a core element in

automated driving vehicles that will be made available in production vehicles by around 2020.

Figure 2.8: Road model based on navigation map information.

The other point of view with respect to map models are digital navigation maps. An example of this type of map is Open Street Maps (OSM), a collaborative project created and updated by a large community around the world. All the information stored in the map is editable and it is freely accessible. The map consists of a list of streets called ways. Every way is composed of a list of nodes with a location and its relations with the other nodes and ways. Thanks to the location and relation between nodes, the shape of the current road and the surrounding streets can be estimated [89], see Figure 2.8.

2.3. Features Integration

All the features described in previous sections are relatively weak in solitary, however, when they work together they can complement each other, obtaining a strong descriptor of the road. The fusion of the features requires some type of optimization process to give relative weights to each feature. Machine learning methods are commonly used for this task. Some examples of those techniques are Support Vector Machine (SVM) [90], Neural Networks (NN) [91], Bayes Classifier, Decision Trees (DT), Random Trees (RT), Extremely Randomized Trees (ERT) and Boosting [30, 92, 93]. They receive a feature vector and the corresponding label for each pixel in the image. After the training stage, the classifier has learned the weight of every feature to the final response. A new classification method was presented in 2012 for

a general purpose classification, however their use for road detection was evaluated in 2014. The name is Convolutional Neural Network (CNN or ConvNet) and it needs high computational requirements. Its complexity requires a Graphical Processing Unit (GPU) for the training stage. One of the most relevant features of this new technique is that instead of receiving a feature vector, it is able to calculate its own feature vector during the training with the image as the unique input. Furthermore, the performance is significantly higher than the other machine learning techniques [94–97]. CNNs always have the problem of overfitting due to lots of connections in the full connection layer. They require a large training set with high variability of scenarios in order to learn all of them.

Machine learning approaches require a post processing stage to smooth their output. Conditional Random Fields (CRF) has been extensively used for this purpose in the literature [98,99]. It learns the response of the classifier for each pixel and also the relation between neighboring pixels in the image. In addition, more complex connections can be created to include temporal relationship between consecutive frames [58].

As reviewed in this section, there are two main type of maps, the first one are navigation maps, which provide information about the steps to reach our destination. The second one are high definition maps, which provide 3D information of the environment with centimeter precision. Most of the autonomous navigation vehicles are based on these type of maps [11]. In contrast to that, our approach is closer to the human way of drive. Human drivers do not need high definition maps. They drive using visual perception and local navigation methods. The only information they need are the indications and steps on the navigation map to reach the destination.

The objective of this thesis is to create a new environment perception method to detect the road in urban environments fusing stereo vision and digital maps by detecting road appearance and road limits

such as lane markings or curbs. CNNs make the system hard coupled to the training set. Even though our approach is based on machine learning (ML) techniques, the features are calculated from different sources (GPS, map, curbs, etc.) and they make our system less dependent on the training set.

2.4. Conclusion

Previous sections have introduced a number of published methods to detect the road using cameras and other type of sensors. Several conclusions can be extracted from it:

- 2D LIDAR sensors provide reliable measurements and make the systems able to detect road limits with accurate precision. However the road limits are only detected at a specific distance. That distance depends on the position and orientation of the sensor. For this reason, measurements should be integrated along the time and ego motion of the vehicle has to be compensated.
- The high number of points and the precision of their 3D coordinates make 3D LIDAR sensors the best sensors to detect the road, curbs or vehicles using their 3D shape for the classification. The drawback of this sensors is the high cost.
- The integration of several 3D LIDAR measurements create a high definition map, which able autonomous vehicles to navigate safely in urban scenarios. The disadvantages of these maps are the high cost of the sensors to integrate the measurements and the 3D LIDAR cost. In addition, the map should be updated frequently to guaranty a safe drive.
- RADAR sensors are a good choice to detect vehicles at long distances in highways and they are able to roughly detect the free space.

- Vision sensors have been used in several proposals due to their low cost and the rich information obtained from an image. The weak point of them is that they are affected by illumination changes provoked by shadows and other weather conditions.
- CNN approaches have demonstrated that are on the top to solve semantic segmentation problems. They are also applied to the road detection problem, aiming robust segmentations. They are strongly dependent of the training dataset, where all the situations in the real scenario should be trained previously.

2.5. Main Contributions

After the review of the state of the art, and considering the discussion presented before, the main contributions of this thesis are:

- The curb detection is an important problem in the context of autonomous vehicles driving in urban scenarios. Most of the analyzed papers use DEM to detect curbs. In this thesis, a new method to detect road curb based on curvature values is presented. The proposed method is able to detect curbs of 3 cm height up to 20 meters long whenever the curb is connected in the curvature image.
- Due to road classifiers based on boosting machine learning techniques cannot manage road limit features such as road markings or curbs, a novel method is developed to integrate those measurements in the classifier. The proposal uses the vanishing point to generate a road model taking information from a digital map and the GPS sensor.
- The method described in the previous point takes information from the digital navigation map and it also provides information of the road width to update the map. It is a kind of symbiosis where

both functions take benefits from the other. The map with the road width is used to generate a prior of the current structure of the road, which is very useful in intersections and narrow streets.

Chapter 3

Development

3.1. Introduction

Autonomous vehicles need to detect the road as an important part of autonomous navigation. This challenging problem is addressed in this thesis with a new method based on computer vision and digital maps.

The computer vision module extracts features based on road texture, color and geometry information. The set of features are collected in a boosting classifier and the outcome is filtered using a Conditional Random Field (CRF) to obtain a smoother result.

The road detection is tackle as a binary classification problem with the labels: road / non road, therefore, boosting classifiers fit well since they are designed for this type of classification.

A public dataset is used in this thesis in order to compare the performance of the proposed algorithm with other state of the art methods. KITTI Vision Benchmark Suite [100] provides images and information of urban scenarios from different sensors, such as monochrome and color cameras, 3D LIDAR, GPS and IMU. The technical characteristics of the sensors used in this thesis are: 2 grayscale cameras 1.4 Mpx

(Point Grey Flea 2), 2 color cameras 1.4 Mpx (Point Grey Flea 2) and Inertial Navigation System (OXTS RT 3003).

Figure 3.1 shows a feature classification depending on the sensor. Greyscale cameras are used to detect road markings, calculate the vanishing point and extract some texture information such as LBP and HOG. Color cameras are important to distinguish vegetation areas that delimit the drivable space in many cases. They are also useful to obtain an illuminant invariant image which is robust to shadows.

Color images can be converted to grayscale, however grayscale cameras have a larger dynamic range and provide images more robust to high illumination contrast.

A pair of grayscale cameras is used as a stereo vision system to estimate 3D information. 3D features are used to detect heights with respect to the ground plane. In addition, big obstacles such as vehicles, buildings or trees are detected using normal vectors and curvatures. Finally, the curvature feature is the base for the curb detection module.

All the features mentioned before are extracted from stereo or monocular cameras. Furthermore, the prior knowledge of the road shape estimated from the navigation map is included. This new feature makes the classifier more robust in situations where camera sensors fail.

In Figure 3.1, the features marked in blue provide high level context information and a deep analysis of them is elaborated in section 3.3. The others marked in red are included in a boosting classifier to detect the free space.

Finally, the potentials estimated during the road classification process are used in a bidimensional Conditional Random Field (CRF) to filter the response of the classifier and obtain a smoother result.

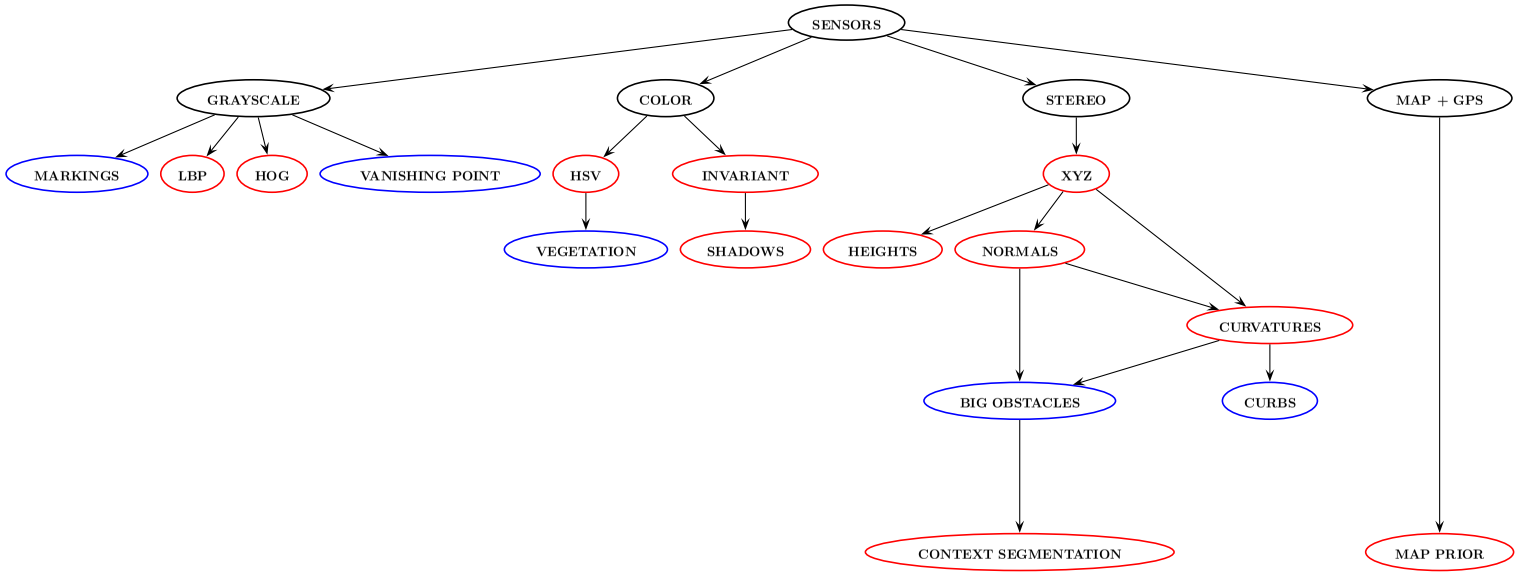


Figure 3.1: Tree of features dependency. The features are classified depending on the source of information: monocular greyscale camera, monocular color camera, stereo greyscale cameras, digital navigation map and GPS. Red nodes mean features included in the boosting classifier and the blue ones inform about the context.

3.2. Features Analysis

Road features are divided in three different types. The first type includes the features that sense road appearance. The second type groups geometry based features and the last one uses a high level set of features since they provide context information.

3.2.1. Appearance Based Features

Appearance features enclose information related with textures and colors. The first ones are analyzed using LBP and HOG and the second ones with HSV, an illuminant invariant image and a shadow detector.

3.2.1.1. Local Binary Pattern (LBP)

Local Binary Patterns (LBP) is a type of visual descriptor that provides texture information of the evaluated pixel using the surrounding values. It has been used in several situations, specially for face recognition or pedestrian detection.

LBP is invariant against any monotonic transformation of the gray scale and can be described by the following formula:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{3.1}$$

$$s(x) = \begin{cases} 1 & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases} \tag{3.2}$$

where P is the size of the neighbor set of pixels in a local neighborhood, R is the radius of the local region, g_c represents the gray value of the center pixel and $g_p(p = 0, 1, \ldots, P-1)$ denotes the gray value of the neighbor.

An example of the LBP codification process in detailed in Figure 3.2, where $g_c = 43$, $g_p = \{60, 20, 40, 32, 52, 10, 24, 80\}$. The resulted

binary pattern is 10001001, which correspond with the value 137 in decimal base. The same procedure is applied in a real urban scenario, resulting the image shown in Figure 3.3.

60	20	40
80	43	32
24	10	52

Figure 3.2: Example of pixel values for LBP codification.

(a) Original image of the scene.

(b) Result of the LBP descriptor.

Figure 3.3: Using a neighborhood of 8 pixels, the resulted descriptor is encoded in a grayscale image.

3.2.1.2. Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients (HOG) is a feature descriptor of the object shape. It has been extensively used for pedestrian detection since 2005 [28]. For this reason the detailed explanation of the algorithm is not included in this section. However, Figure 3.4 shows the process to obtain the HOG descriptor. The structured shape of common road limits (curbs, lines or other type of obstacles) can be characterized by the HOG descriptor. Our specific configuration creates windows or blocks of 16 pixels, with cells of 8 pixels and the histogram is computed with 9 bins, obtaining a vector of 36 elements.

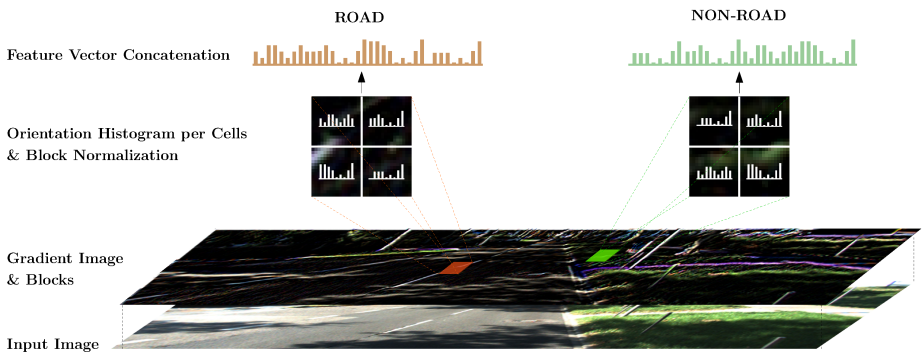


Figure 3.4: HOG descriptor estimation.

3.2.1.3. HSV

Color cameras usually create images coded with RGB color space. The perceived difference of color between two points in the RGB color space is not equal to its distance and their channels are hardly correlated. Contrary to RGB cartesian representation, the polar 3D coordinates of HSV create a color uniform distribution that makes object segmentation more intuitive.

3.2.1.4. Illuminant Invariant Image & Shadow Detection

Road detection using computer vision is a challenging task, specially when the road is affected by shadows. The idea of many computer vision approaches is to consider that the road have some constant features like color or texture that can be grouped. Road texture and color sometimes are not homogeneous because the asphalt has an irregular degradation or it has been partially renewed. Even in homogeneous textured roads, the road is strongly affected by shadows, creating a new challenging scenario for computer vision algorithms.

Some approaches try to attenuate the shadow influence by an illuminant invariant space to detect the road [101], [37]. As explained in

[36] and [102], adopting certain assumptions about lights and cameras, color images can be represented in a 1D shadow free grayscale image.

The steps to obtain an illuminant invariant image start with the creation of a two-vector chromaticities χ:

$$\chi_j = \frac{\rho_k}{\rho_p}, k \in \mathbb{Z}_3, k \neq p, j \in \mathbb{Z}_2 \tag{3.3}$$

For an RGB image, $p = 2$ means $\rho_p = G, \chi_1 = R/G, \chi_2 = B/G$. The chosen channel ρ_p makes the algorithm unstable because the result depend on the dominant color of the scene. A better solution is to divide by geometric mean and do not favor one particular channel, see equation 3.4.

$$\chi_j = \frac{\rho_k}{\sqrt[3]{\prod_{i=1}^{3} \rho_i}} \tag{3.4}$$

The logarithm of the chromaticities $\boldsymbol{\chi}' = log(\boldsymbol{\chi})$ is represented in Figure 3.5. Given this representation, as illumination color changes, the log-chromaticity vector $\boldsymbol{\chi}'$ for a given surface moves along a straight line and the direction ($\mathbf{e}$) of this line depends on the properties of the camera. By projecting the log-chromaticity vector $\boldsymbol{\chi}'$ onto the vector orthogonal to $\mathbf{e}$, which is denoted as $\mathbf{e}^{\perp}$, a 1D illuminant invariant representation can be obtained following equations 3.5 and 3.6.

$$\mathcal{I}' = \chi' \mathbf{e}^{\perp} \tag{3.5}$$

$$\mathcal{I} = exp(\mathcal{I}') \tag{3.6}$$

The resulted illuminant invariant image is shown in Figure 3.7b. As mentioned before, the direction of $\mathbf{e}$ should be calibrated because the angle θ depends on the camera. The intrinsic parameter calibration is an offline process and it can be done using two main approaches: The first one was presented in [103] and uses a Macbeth color checker

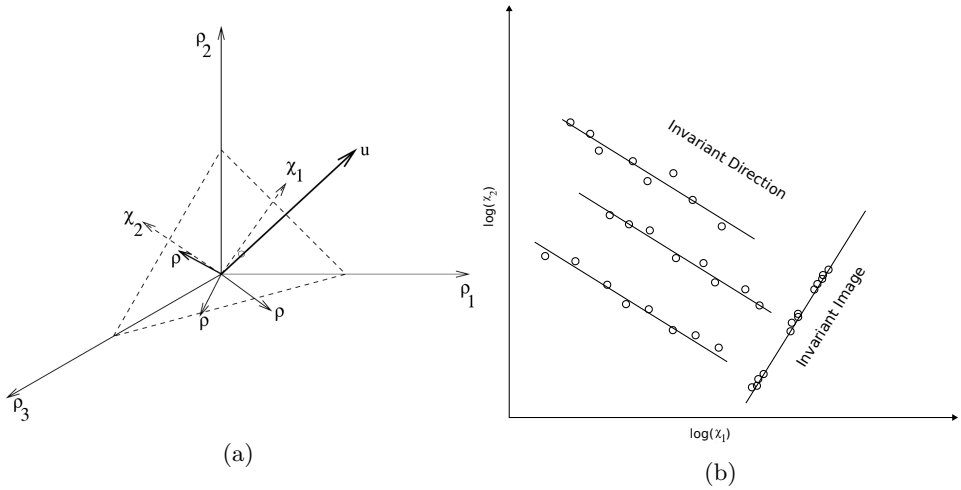


Figure 3.5: Figure 3.5a shows the representation of the 2D chromaticity space. Figure 3.5b shows the representation of an ideal camera and Planckian illumination. The chromaticities move along a straight line with and specific direction, which depends on the camera.

under different daytime illuminations to obtain θ. The second approach [104], analyses the histograms of the invariant image. The direction α that generates an invariant image with minimum entropy is the correct angle. The entropy is calculated following equation 3.7, where n is the number of bins of the histogram H_α.

$$\eta_\alpha = -\sum_{i}^{n} H_i log(H_i) \tag{3.7}$$

In order to obtain the minimum entropy, the log-chromaticity ρ is projected on a direction α from 0 to 180 degrees. Figure 3.6 shows the minimization process and how the correct angle creates an illuminant invariant image. The second approach presents some advantages over the first one. The first one is that the calibration is based on regular images, that means no special pattern images are required. Another one is that it can be computed over several images and get a more robust calibration. For these reasons, the second approach is the one implemented in our system.

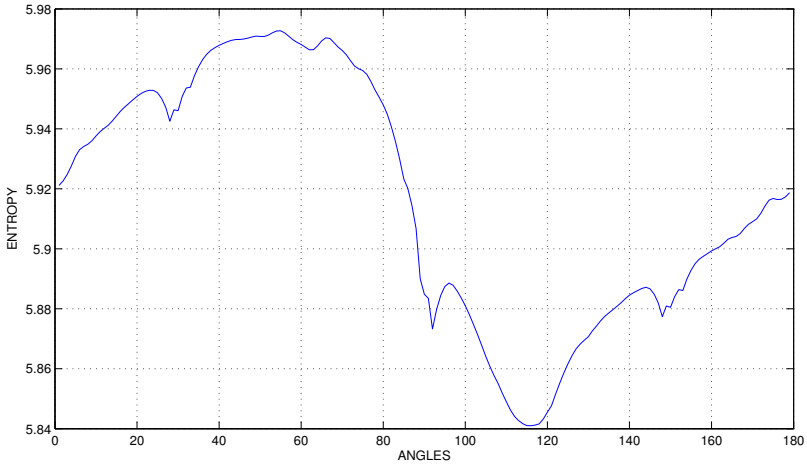


Figure 3.6: Representation of the entropy minimization for camera calibration.

Consider again the 2D chromaticity representation defined in equation 3.5. In this representation, illumination is represented by a vector of arbitrary magnitude in the direction **e**:

$$illumination = \chi'_E = a_E e \tag{3.8}$$

This magnitude indicates the illumination of the scene, which is a very good feature to obtain the shadows of the scene, see Figure 3.7c. When the shadows of the scene are detected, a wide range of possible applications appear:

- It can be a new feature for a classification method.
- It can be used as a mask for a special image processing method on shadowed areas.
- It can be used to obtain a new shadow free image.

This 1D representation can be extended to a 2D chromaticity representation. In this 2D representation, it is possible to relight all pixels to a shadow free image. Finally, a 3D full color shadow free image can

be reconstructed detecting shadowed areas filling their color with the surrounding pixels.

(a) Original image of the scene.

(b) Illuminant invariant image.

(c) Illumination of the scene.

(d) Shadow detection.

Figure 3.7: Results of illuminant invariant and shadow detection images.

Given a shadowed region S, it is necessary to detect the area S_{road} that belong to the road and also the other that belong to non road surface $S_{\overline{road}}$. In addition, it is required to detect the lighted areas of road and non road surfaces to fill the shadowed pixels with the corresponding lighted values. It creates a new challenging problem. For that reason, the shadow free image is not developed for our road classifier.

3.2.2. Geometry Based Features

Objects shape is an important characteristic for their classification. Road is nearly an horizontal flat surface and vertical obstacles are clearly separable using geometry descriptors.

3.2.2.1. XYZ

In the field of computer vision, a pair of stereo cameras is able to create a 3D representation of the scene. The factors that influence in the reconstruction are the focal length and the distance between the cameras. As mentioned in section 3.1, the KITTI dataset is configured with a pair of cameras separated 0.54 meters at 1.73 meters with respect to the road surface and equipped with 8 mm focal length optics. After rectification, the image size is 1242 width and 375 pixels height.

Using the semi global block matching (SGM) algorithm to estimate the disparity map, the system generate a dense 3D point cloud where the closer points are at 6 meters. As demonstrated in following sections, the incorrect disparity values have a direct impact in the response of the perception methods.

3.2.2.2. Normal Vectors And Curvature Variation

In urban scenarios the road is usually limited by curbs, road markings or pavement with a different texture. Depending on the city, curbs have a wide range of heights. For this reason, it is necessary to create an adaptive method to detect every curb, regardless the curb height.

Our approach is focused on the curvature feature. This feature describes a local surface variation and it was applied for 3D semantic perception in [105]. The input is an unstructured 3D point cloud $P = \{\mathbf{p}_i \in \mathbb{R}^3\}$. The local features are computed on the nearest neighbors N_p enclosed in a sphere of radius r around the sample p, thus $r \geq \max_{i \in N_p} \|\mathbf{p} - \mathbf{p}_i\|$. As demonstrated in [106] and [107] eigen-analysis of

the covariance matrix of a local neighborhood can be used to estimate local surface properties. The 3×3 covariance matrix C for a sample point $\mathbf{p}$ is given by equation 3.9, where $\bar{\mathbf{p}}$ is the centroid of the neighbors $\mathbf{p}_{i_j}$ of $\mathbf{p}$.

$$C = \begin{bmatrix} \mathbf{p}_{i_1} - \bar{\mathbf{p}} \\ \vdots \\ \mathbf{p}_{i_k} - \bar{\mathbf{p}} \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{p}_{i_1} - \bar{\mathbf{p}} \\ \vdots \\ \mathbf{p}_{i_k} - \bar{\mathbf{p}} \end{bmatrix}, i_j \in N_p \tag{3.9}$$

Since C is symmetric and positive semi-definite, all eigenvalues λ_l are real-valued. Corresponding to the principal components (PCA) of the point set defined by N_p, the eigenvectors $\mathbf{v}_l$ form an orthogonal frame. The eigenvalues $\lambda_l \forall l \in \mathbb{Z}_3$ measure the variation of the $\mathbf{p}_i \in N_p$ along the direction of the corresponding eigenvectors.

In order to compute the surface curvature, a tangent plane to N_p is required. The plane $T(\mathbf{x})$ is represented as a point $\mathbf{x}$ and a normal vector $\mathbf{n}$. Assuming $\lambda_0 \leq \lambda_1 \leq \lambda_2$, the plane satisfies equation 3.10.

$$T(\mathbf{x}) : (\mathbf{x} - \bar{\mathbf{p}}) \cdot \mathbf{v}_0 = 0 \tag{3.10}$$

Through $\bar{\mathbf{p}}$ minimizes the sum of squared distances to the neighbors of $\mathbf{p}$. Therefore $\bar{\mathbf{v}}_l$ approximates the surface normal $\mathbf{n}_p$ at $\mathbf{p}$. λ_l describes the variation along the surface normal, that means how much the points deviate from the tangent plane. Finally, the surface curvature γ_l at point $\mathbf{p}$ is defined as shown in equation 3.11.

$$\gamma_l(\mathbf{p}) = \frac{\lambda_l}{\sum \lambda_l} \tag{3.11}$$

After the normalization, the curvature values vary between 0 and 1, where low values correspond to flat surfaces. The result is a vector $\boldsymbol{\gamma}$ similar to surface normal vectors, moreover curvature vectors are more stable and robust. The objective is to detect variations on the road surface, for this reason we only take into account the component

γ_z, which is orthogonal to the road plane in our reference system, see Figure 3.8.

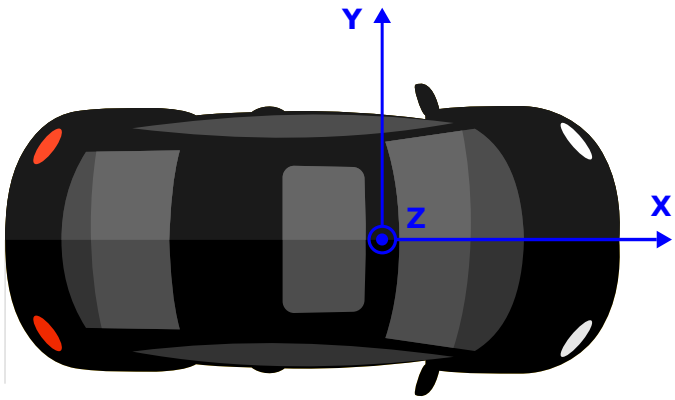


Figure 3.8: Reference system of the ego vehicle.

The curvature variation is computed on an artificial point cloud to show the responsiveness of the algorithm. The point cloud has several curbs of different heights, starting with 3 cm in steps of 2 cm up to 15 cm. As shown in Figure 3.9 the feature provides enough information to detect variations of the curvature even for a curb of 3 cm height.

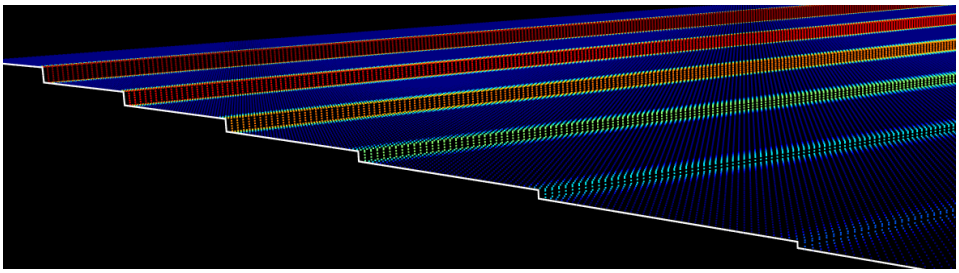
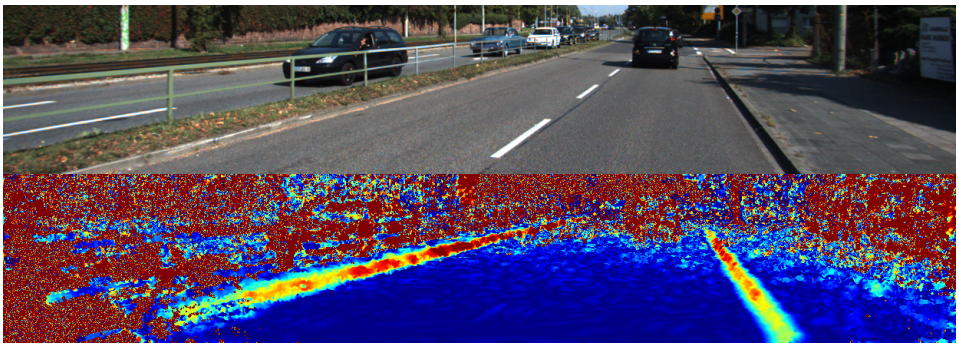


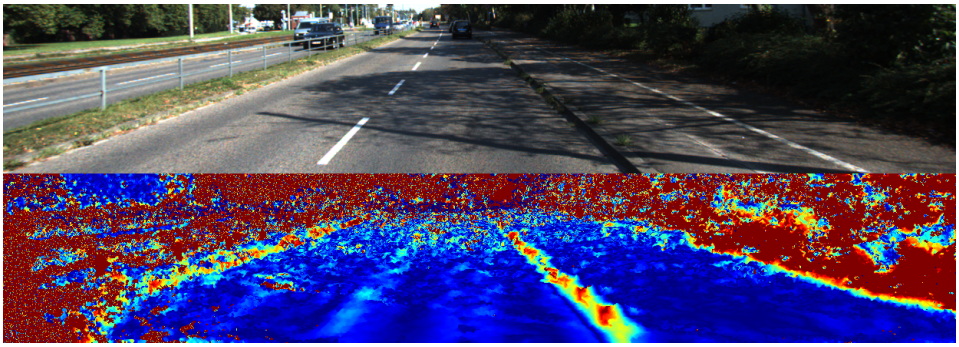
Figure 3.9: Curvatures of an artificial 3D point cloud with curbs of different heights: 3, 5, 7, 10, 12 and 15 cm respectively. Curvature values are represented in a color scale where cold colors correspond to low curvature values and warm colors correspond to high values.

Real scenarios differ from the ideal curvature estimation shown in Figure 3.9. In Figure 3.10, a set of real scenarios is presented, where mismatching errors during the stereo computation provoke invalid curvature values (see discontinuous red areas).

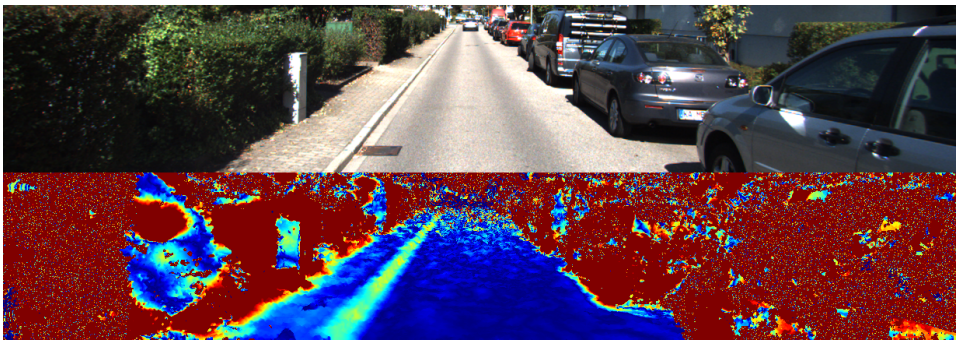
Due to the fact that the system relies on stereo vision, curvatures are robust to illumination changes (Figure 3.10b). Residential areas are a challenging scenario to detect road limits because curbs are small and their detection is very important to drive safely. Curvature variations are visible even for small curbs. Figures 3.10c, 3.10d and 3.10e demonstrate that this feature is able to distinguish curbs of 3 cm even at far distances.



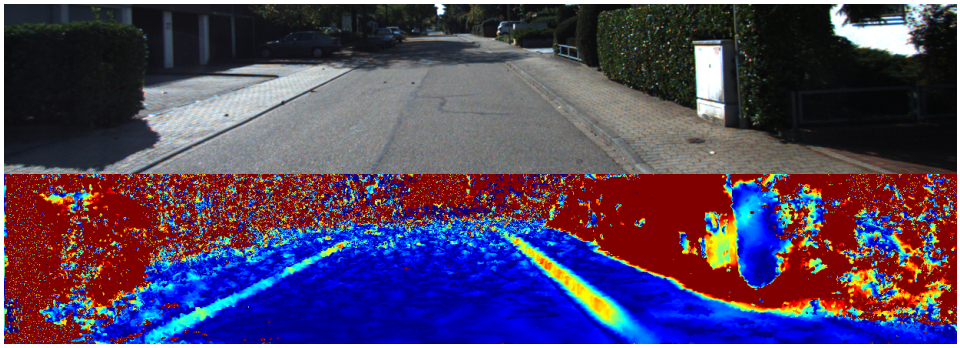
(a) Big curbs on urban road.



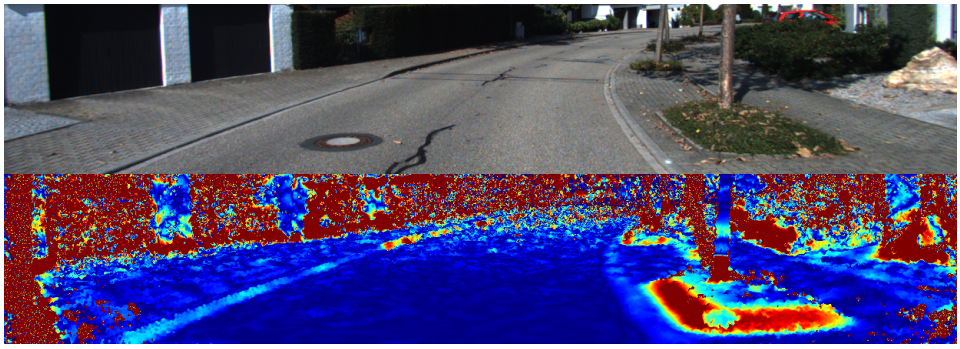
(b) Robust to shadows.



(c) Residential scene.



(d) Small and regular curbs on residential scenes.



(e) Small curbs.

Figure 3.10: Curvature values on different urban scenes.

3.2.2.3. Heights With Respect To The Ground Plane

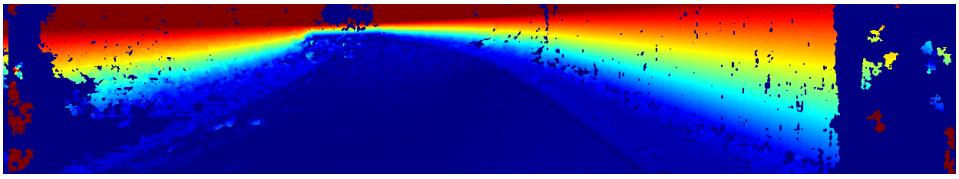
The road shape can be approximated to a plane in some cases. Given that vertical obstacles have large distances with respect to the ground, the smaller distance to the plane, the most probable to belong to the road.

The first step is to estimate the ground plane. It is performed using RANSAC on the 3D point cloud obtained from the stereo cameras. Given a plane $ax + by + cz + d = 0$ and a point $\boldsymbol{x_0} = (x_0, y_0, z_0)$, the normal vector to the plane is given by $\boldsymbol{v} = \begin{bmatrix} a, b, c \end{bmatrix}^T$ and a vector from the plane to the point is given by $\boldsymbol{w} = - \begin{bmatrix} x - x_0, y - y_0, z - z_0 \end{bmatrix}^T$. Equation 3.12 projects $\boldsymbol{w}$ onto $\boldsymbol{v}$, giving the distance D from the point to the plane.

$$D = \frac{|\boldsymbol{v} \cdot \boldsymbol{w}|}{|\boldsymbol{v}|} \tag{3.12}$$

Points at 3 cm of less from the estimated plane are considered inliers, see Figure 3.11a. Then, distances between the plane and every pixel in the image are computed. The new feature is shown in Figure 3.11b, where cold colors mean small distance to the ground plane.

(a) Plane inliers.



(b) Heights encoded in a color scale.

(c) Wrong ground plane estimation.

Figure 3.11: Heights with respect to the ground plane.

This feature is useful whenever the ground plane is well detected. In Figure 3.11c, an example of a wrong road plane estimation is depicted. In order to improve plane approximations, the searching area can be restricted discarding further pixels.

3.2.3. Context Based Features

Some features produce higher level information than others. The ones detailed in this section are not relevant to describe the road. However, they inform about the context of the scene, which is very useful to understand how the road is distributed.

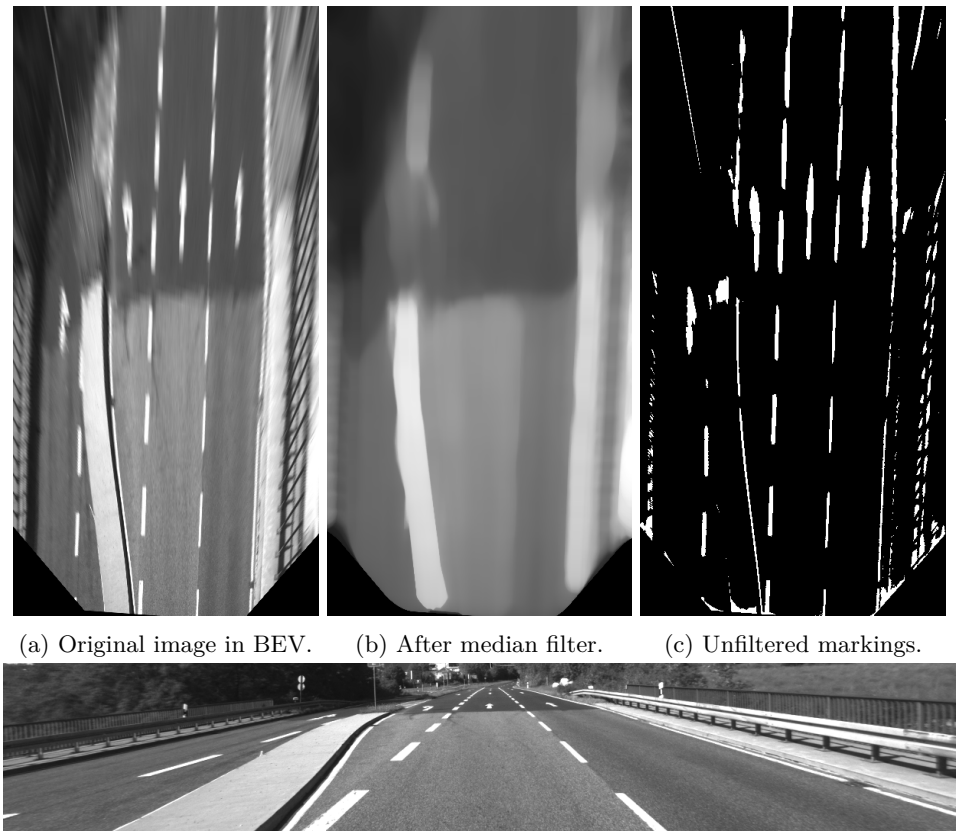
3.2.3.1. Road Markings

Road markings detection is a basic task for autonomous navigation. In a multi-lane scenario the free space has to be split in lanes and road markings are crucial for this task. The road markings are not always present and sometimes they are partially or totally removed. However, when the road has had a correct maintenance, they provide relevant information about the road limits and traffic rules.

This topic has been extensively studied in the last decades. For this reason, the proposed road marking detection method is based on state of the art techniques. Nevertheless, we provide a brief description for completeness purpose.

As explained in [47], a median filter is applied to the input image. The window size should be adjusted due to perspective. In order to keep the window size constant, a bird eye view (BEV) of the scene is reconstructed. Lane markings appear parallel in the new view because they are not distorted by the perspective.

The window size of the median filter needs to be twice larger than the road marking, otherwise the border is well detected but the areas inside the zebra crossing are not. The median filter image is subtracted to the original one and the result is thresholded using an adaptive threshold. Finally, contours are filtered by shape and size to remove some noise.



(a) Original image in BEV. (b) After median filter. (c) Unfiltered markings.

(d) Original image of the scene.

Figure 3.12: Process of road markings detection.

3.2.3.2. Vegetation

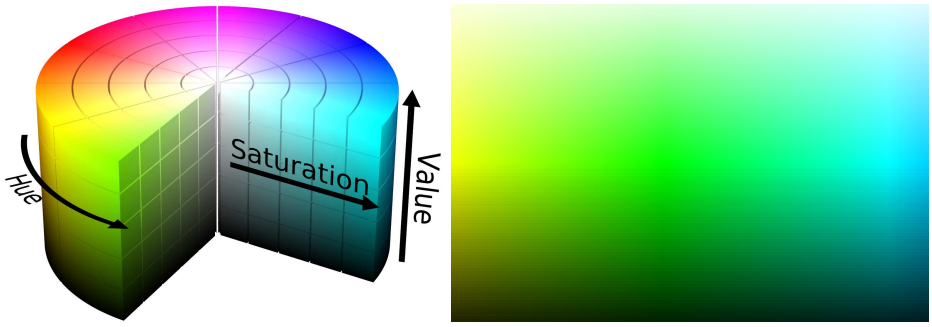
The use of cameras for the environment perception provides information of texture and color. The texture information was exploited in previous sections to estimate LBP descriptors or to detect road markings. In this section, color information is utilized to detect vegetation areas.

Despite the fact that the road surface has a wide range of textures and color values, green areas usually correspond with vegetation areas in the majority of scenarios. For this reason, the free space detection system considers these areas as not drivables.

The RGB color space is not a good choice to detect vegetation because the Red, Green and Blue channels are strongly correlated. Furthermore, the color difference between two colors is not linearly dependent on the distance between them. For this reason a transformation to Hue Saturation Value (HSV) color space is performed. This color space splits the dominant color (H), the saturation of the color (S) and the brightness (V), see Figure 3.13a.

By selecting a range of hue values, all the variations of green are detected. This color selection requires another two restrictions in saturation and brightness channels. Small saturation values mean a color close to white, and small brightness values mean a color close to black. In order to reduce the number of false positives in white (road markings, white vehicles) and dark areas (shadowed roads, black vehicles), a minimum value is applied to saturation and brightness channels. The color selection is shown in Figure 3.13b and it is accomplished following equation 3.13.

$$\begin{aligned} \mathbb{G}_h &= H[\alpha_0, \alpha_1] \\ \mathbb{G}_s &= S[\beta_0, \beta_1] \\ \mathbb{G}_v &= V[\gamma_0, \gamma_1] \\ \mathbf{p} \in \mathbb{G} \iff p_h \in \mathbb{G}_h \wedge p_s \in \mathbb{G}_s \wedge p_v \in \mathbb{G}_v \end{aligned} \tag{3.13}$$



(a) HSV color space. (b) Colors of interest.

Figure 3.13: Colors of interest in HSV color space for vegetation detection.

The color segmentation applied to real scenarios fail sometimes in presence of shadowed areas, see Figure 3.14. For this reason, morphological operations are used to partially fill misdetections. In our system it is preferred to misdetect some dark green areas and keep a low false positive rate.

(a) Original image of the scene.

(b) Result of the vegetation detection.

Figure 3.14: Green areas detected using HSV color space.

3.2.3.3. Road Curbs

As explained in section 3.2.2.2, the curvature variation is a good feature to detect curbs, nevertheless in a real scenario the curvature values have not homogeneous values due to mismatching errors during the stereo disparity map computation, see Figure 3.10. Depending on the curb height, the curvature values are different in each scene. The challenge is to detect every curb using the same algorithm and regardless the curb height.

The details of the algorithm are explained as follows. In our reference system, the Z axis is orthogonal to the road plane, therefore the curvature γ_z provides a discriminative descriptor of the road shape. Curb height and curvature γ_z are highly correlated. Consequently, after thorough observation of urban scenes, a set of thresholds

$\alpha_i = \{1 \dots N\}$ is used to label the type of curb, see Table 3.1.

Table 3.1: Curb Curvature Values

DESCRIPTION	CURVATURE	COLOR
Flat surface	$0 \leq \gamma_z < \alpha_0$	not painted
Very Small Curbs (∼3 cm)	$\alpha_0 \leq \gamma_z < \alpha_1$	yellow
Small Curbs (∼5 cm)	$\alpha_1 \leq \gamma_z < \alpha_2$	orange
Regular Curbs (∼10 cm)	$\alpha_2 \leq \gamma_z < \alpha_3$	red
Big Obstacles	$\alpha_3 \leq \gamma_z \leq 1$	purple

The thresholds selection is calculated offline, classifying curvatures in 5 groups. The first one collects samples that lie on a flat surface. This group will be considered as a flat surface and therefore this group will be filtered and not taken into account for the curb detection. The second one has a range that corresponds with very small curbs, that is the ones close to 3 cm height. The third cluster includes small curbs of 5 cm height and finally the regular curbs ($\sim$ 10 cm) are grouped in the fourth group. Finally, the other samples with larger curvatures will be considered as big obstacles. In this group will be included cars, pedestrians, traffic signs, fences, walls, etc. and are not considered during the filtering process.

The resulted clusters are filtered independently using morphological operations and contour analysis because the image has several noisy measurements. The filtered clusters are merged back and the new clusters are considered as road curbs. Figure 3.15 shows all the process in a block diagram.

This approach provides flexibility because the algorithm works on a wide range of scenes and for every type of obstacles, detecting in every case the most dominant curvature value. The ones from 3 cm height are detected as well as the others with larger heights.

The use of fixed or empirical thresholds is then avoided given that the proposed function is adapted automatically for different scenes depending on the predominant curvature value. For example, if the

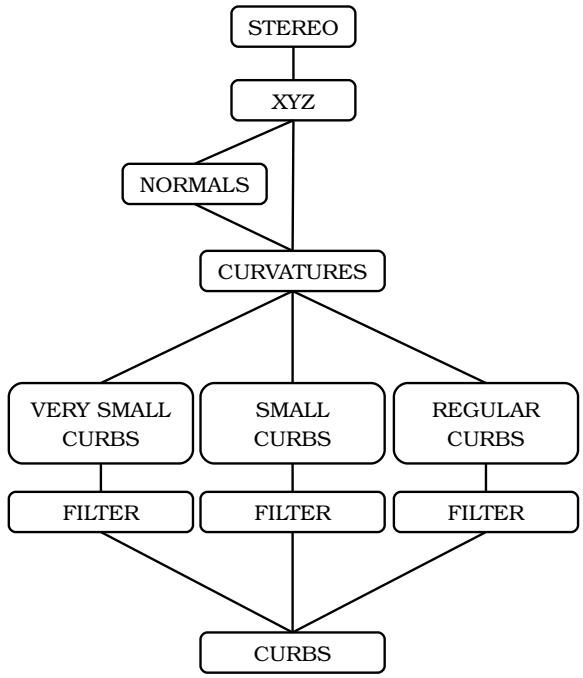


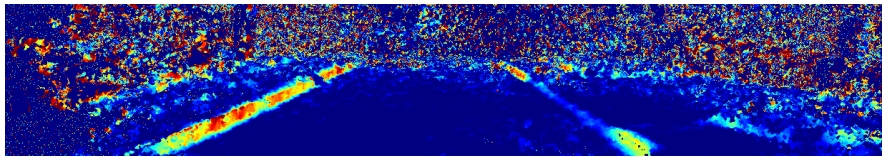
Figure 3.15: Diagram of curb detection algorithm.

curb is a regular one, most of the points exhibit curvature values $\gamma_z \in [\alpha_2, \alpha_3)$, but there are also some curb points yielding significantly different values. The method is based on 3D information, therefore it is robust to different illumination conditions and to different textured roads, however, the performance is affected by the mismatching errors.

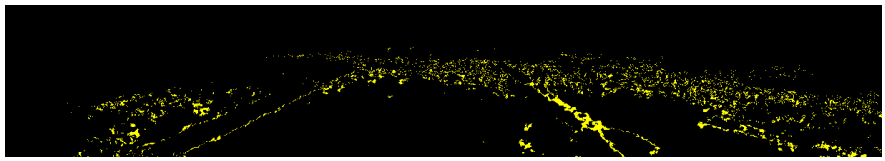
Figure 3.16 shows a residential area with curbs of different heights. On the left, a regular curb separates the road and vegetation areas. On the right side, a regular curb is combined with a very small curb due to the access to a parking area. The heights of both curbs are very different since the big one is $\sim$10 cm and the small one is $\sim$3 cm.

The analyzed scene demonstrates the versatility and robustness of the presented algorithm. Whenever the curb is connected in the curvature image, the very small curbs are correctly detected up to 20 meters and the method is robust to a wide range of curb heights without any parameter adjustment.

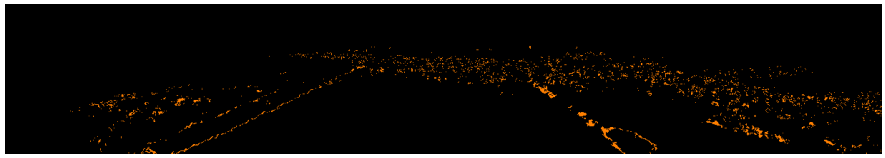
(a) Input image.



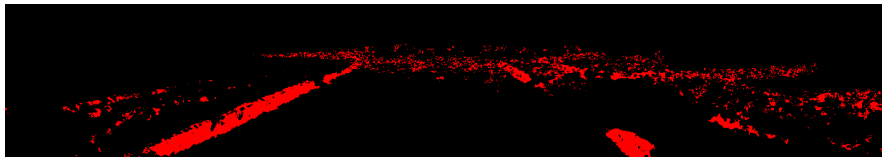
(b) Unfiltered curvature values.



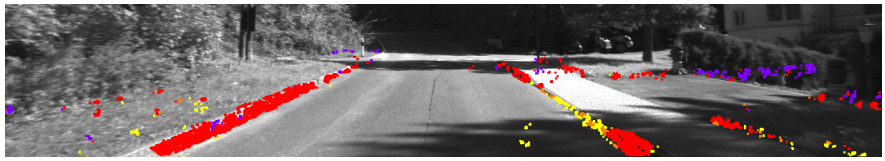
(c) Unfiltered mask for $\gamma_z \in [\alpha_0, \alpha_1)$



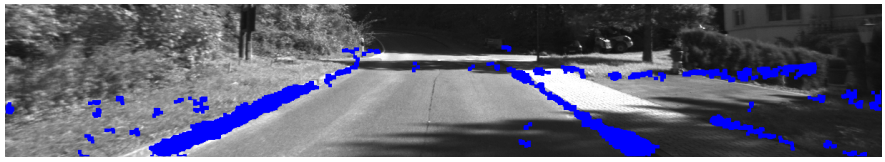
(d) Unfiltered mask for $\gamma_z \in [\alpha_1, \alpha_2)$



(e) Unfiltered mask for $\gamma_z \in [\alpha_2, \alpha_3)$



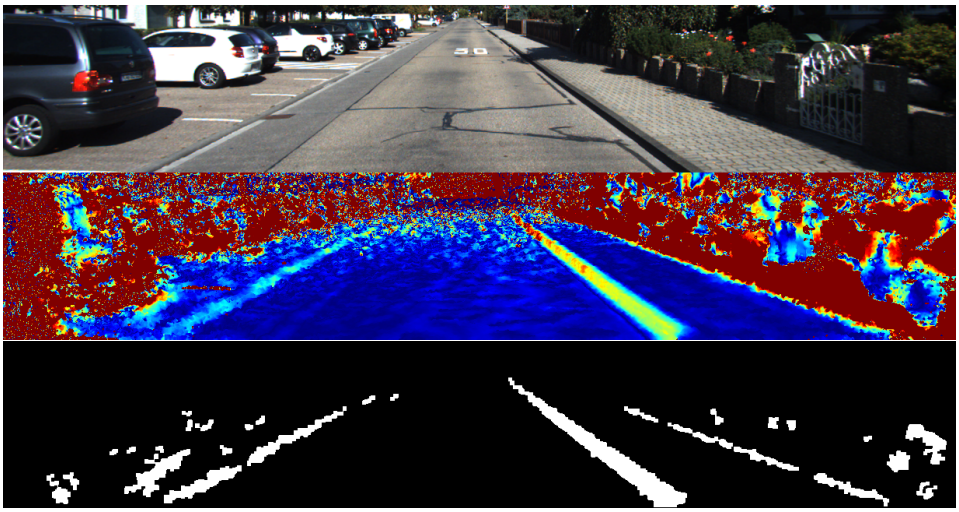
(f) Filtered clusters.



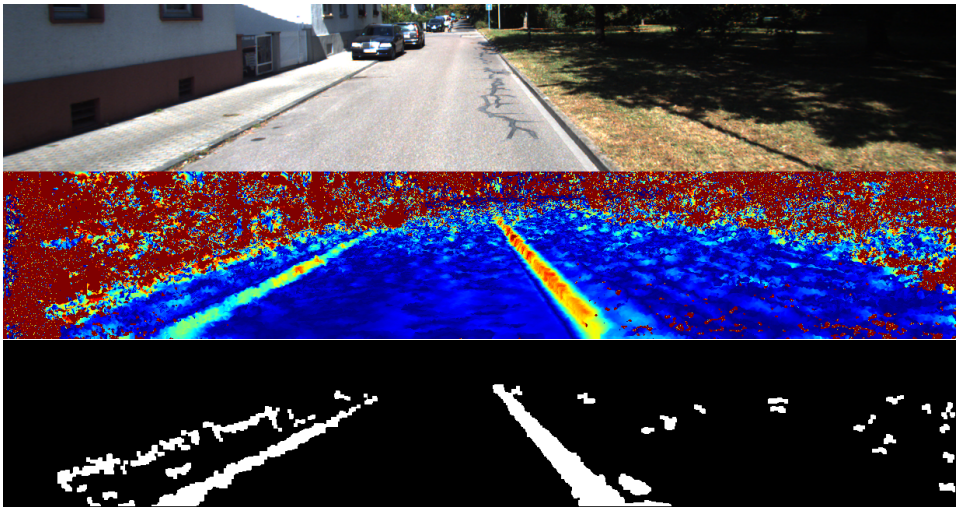
(g) Final result of road curb detection.

Figure 3.16: Progression of the road curb detection algorithm.

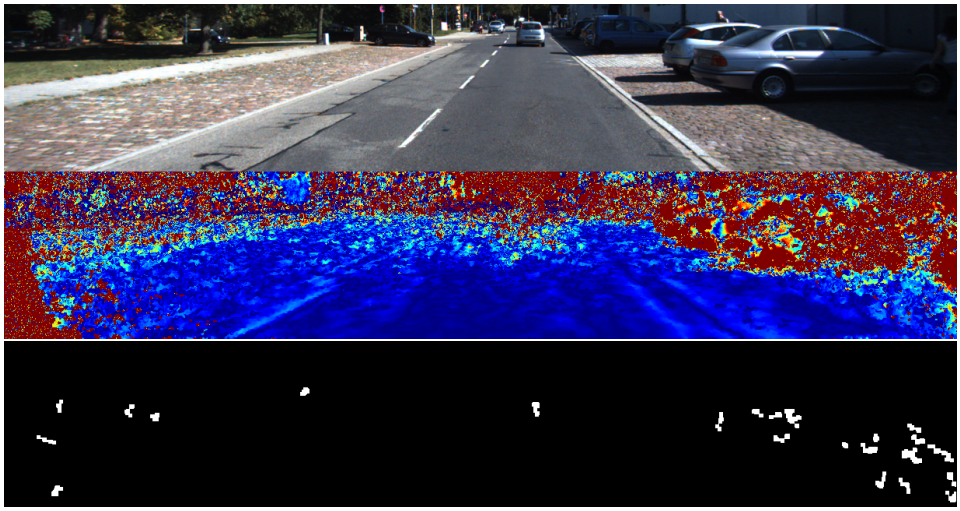
Figure 3.17 shows the results obtained in different scenarios. Figures 3.17a and 3.17b are residential areas. In the first one there is a sidewalk on the right and a parking place on the left. In the second one, the road is limited by grass on the right and a small curb on the right. Those scenes remark the importance of context information to understand the structure of the scene properly. The method has limitations regarding to stereo matching and minimum height required. Figure 3.17c demonstrates that geometry is not discriminative enough



(a) Residential scene with a parking area separated by a small curb.



(b) Residential scene with a small curb on the left and a regular one on the right.



(c) Challenging scenario where the system is not able to detect the small curb.

Figure 3.17: Curb detection in different scenarios.

in some situations and it is necessary to include complementary features such as texture or color.

3.2.3.4. Big Obstacles

The free space is usually limited by curbs, road markings, vegetation areas or other obstacles, such as buildings, parked cars, post lamps, traffic lights or traffic signs. This type of obstacles are detected using 3D information from the stereo cameras.

A general description of the algorithm is shown in Figure 3.18. The 3D points are processed to estimate normal and curvature vectors as mentioned in section 3.2.2.2. The points with components $n_x \geq 0.5$ or $n_y \geq 0.5$ or $\gamma_z \geq 0.5$ are considered big obstacles. Some pixels have noisy or unrealistic vector values. For this reason, every component is filtered by area independently. Afterwards, they are merged to obtain the final result.

Qualitative results are shown in Figure 3.19, where segmentation with filtering improve the segmentation without the filtering.

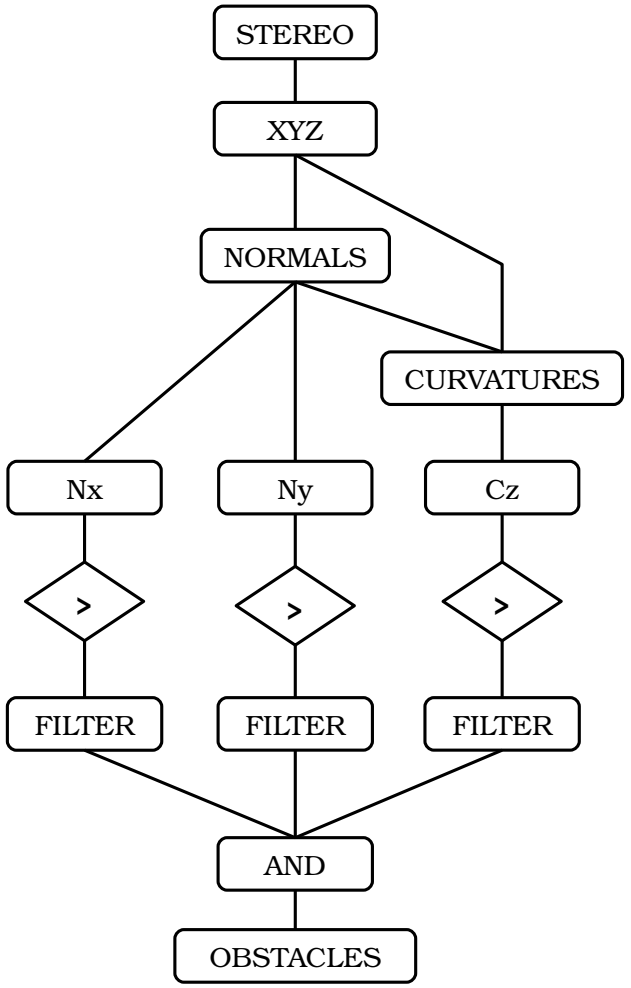
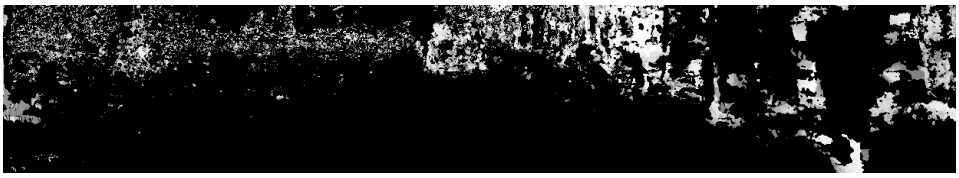


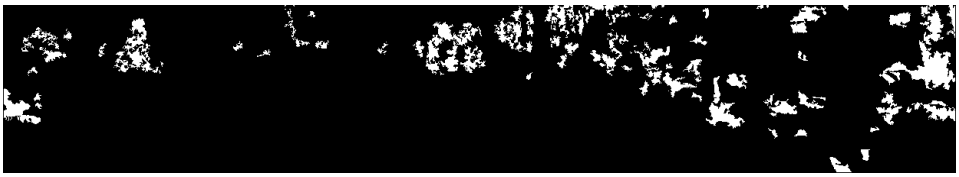
Figure 3.18: Diagram of obstacle detection algorithm.

Big obstacles are in most of the cases vertical obstacles such as vehicles, post lamps, trees, traffic signs, pedestrians, buildings, etc. The holes depicted in Figure 3.19c are caused by mismatching errors. In order to get a more robust result, columns with an obstacle are considered occupied from the bottom row of the obstacle until the first row of the image. The resulted image has obstacles without holes inside, obtaining a more realistic representation of the scene, see Figure 3.19d. The drawback of this approach is that small false positives detections create larger false positives areas.

(a) Original image of the scene.



(b) Big obstacles detection without noise filter.



(c) Big obstacles detection with noise filter.

(d) Final result with noise filter and vertical objects projection.

Figure 3.19: Big obstacles detection process. Vertical projection of the obstacles make the feature more realistic to the scene.

Finally, Figure 3.20 collects different scenarios where the obstacles detection has been tested. In particular, Figure 3.20c shows how some segments that belong to the curb are detected as big obstacles due to the size of the curb. As demonstrated in Figure 3.20f, the errors during the disparity image create some false positive detections on the left of the image.

(a) Scene with a wall on the right and a cyclist riding close to the vehicle.

(b) Correct detection of parked vehicles.

(c) Correct detection of vehicles, a bush on the right and a group of cyclist standing on the sidewalk.

(d) Curbs detected as obstacles due to their size.

(e) Far obstacles are also correctly detected.

(f) Correct detection of traffic signs in the closer distance but incorrect detection of obstacles on the left.

Figure 3.20: Final result of the obstacle detection method in different scenarios.

3.3. Road Segmentation Based On Context Information

The method described in section 3.2.3.4 demonstrates that the free space can be roughly detected with it, however, curbs, road markings and vegetation areas are also elements that restrict the drivable area. Those features require a high level algorithm to understand how the free space is distributed along the image.

There are two main types of features, the first type describes the road, and they can be included directly to the classifier. Some examples of these features are normal vectors, 3D points, HOG, LBP, etc. The second type of features are road limits, such as road markings, curbs, vegetation areas or obstacles. These features provide important information of the scene, however they should not be included directly to the classifier (curbs and road markings) because the features do not describe the road neither the non road area.

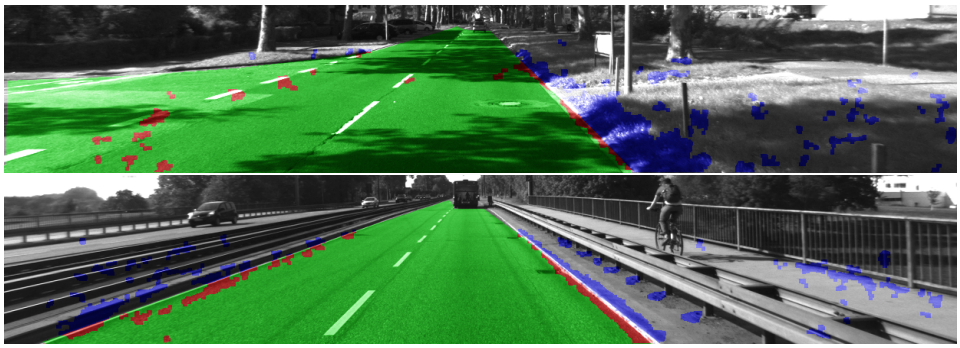


Figure 3.21: The road curb is not good feature for a road/non road classifier because half of the measurements are positive values of road and the others not.

Figure 3.21 shows some examples of curb detection combined with the ground truth of the road (green). The pixels in red correspond with the curb points that lie on the road and the blue pixels are the ones that lie out of the road. That is the reason why in [30], the weights that correspond to those features were very small in the final classification response. In order to increase the weight of those features,

the proposed method converts from limit features to a new feature that describe the road.

Some approaches in the state of the art use virtual rays starting from the bottom of the image to detect the road boundary. Analyzing the feature values along the rays, the point that satisfies some conditions is established as the road limit for that ray. The connection of all of them creates a closed polygon that is considered free space.

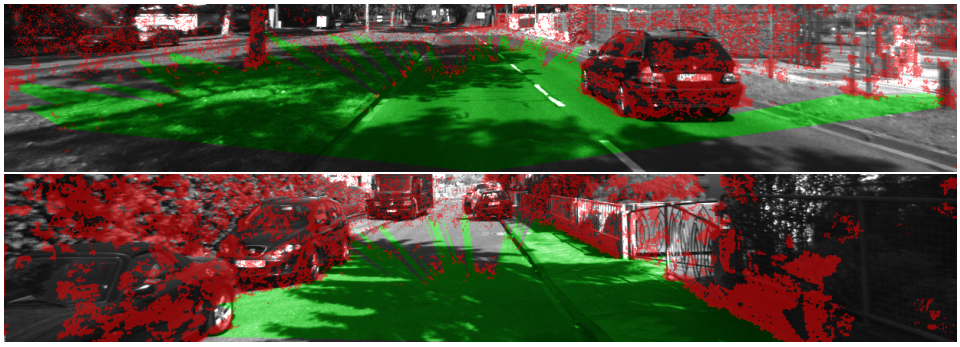


Figure 3.22: Example of free space detection using a set of rays starting from the bottom center of the image.

As demonstrated in Figure 3.22, this point of view is sensitive to noisy measurements. In this thesis, the proposed analysis relies on the vanishing point as starting point of the rays. This point of view is more robust to wrong measurements and it is more intuitive.

Regarding the pin hole camera model, straight parallel lines converge in a vanishing point. The road limits are in most of the cases parallel, even in narrow roads. For this reason, a set of rays are estimated starting from the vanishing points.

First of all, the vanishing point has to be detected. There are several published papers related with this topic. Since the goal in this thesis is not the development of a novel method to obtain vanishing points, a public library has been used for this purpose [108].

A brief description for completeness purpose is provided. The first step is the detection of edges in the image using the Canny edge detec-

tor. The second one is to find straight lines using the Hough transform and finally the vanishing point is estimated using M-estimator SAmple and Consensus (M-SAC).

RANSAC can be sensitive to the choice of the correct noise threshold that defines which data points fit a model. If such threshold is too large, then all the hypotheses tend to be ranked equally. On the other hand, when the noise threshold is too small, the estimated parameters tend to be unstable. To partially compensate this undesirable effect, the M-SAC [109] variation tries to evaluate the quality of the consensus set calculating its likelihood.

The general description of the method is explained in Figure 3.23. The main goal of the algorithm is to find the radial rays that can be the road limits. Firstly, a set of radial rays from the vanishing point are analyzed along the image. The sum of the features along the ray are displayed in Figure 3.24a.

After smoothing, the first derivative is calculated. Curbs or road markings features create two strong symmetric peaks on the feature first derivative. Supported on Bolzano's theorem, the angles γ that satisfies the equation 3.14 are included to a rays vector.

Theorem 3.3.1. *Bolzano's theorem states that if f is a continuous function in the closed interval $[a, b]$ with $f(a)$ and $f(b)$ of opposite sign, then there is a c in the open interval (a, b) such that $f(c) = 0$.*

$$\begin{aligned} |f'(\alpha)| &\geq th_c \\ |f'(\beta)| &\geq th_c \\ |f'(\gamma)| &= 0 \\ sgn(f'(\alpha)) &\neq sgn(f'(\beta)) \\ \forall\ \gamma &\in [\alpha, \beta] \end{aligned} \tag{3.14}$$

The obstacles and vegetation have a different distribution, they only create a single peak on the feature first derivative since their values

are not symmetric with respect to the road limit. Vegetation areas are usually big surfaces out of the road. The false positives of the

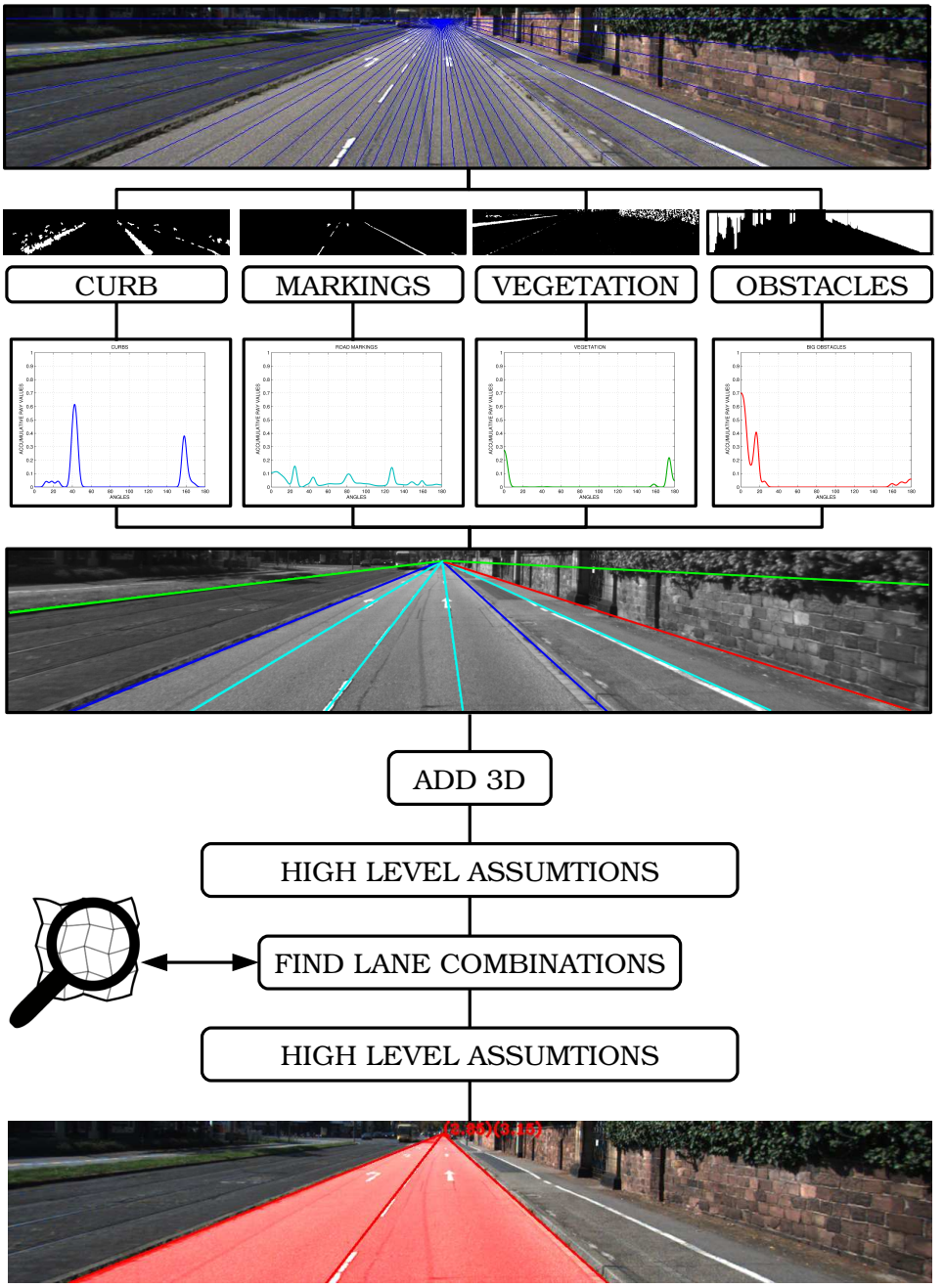
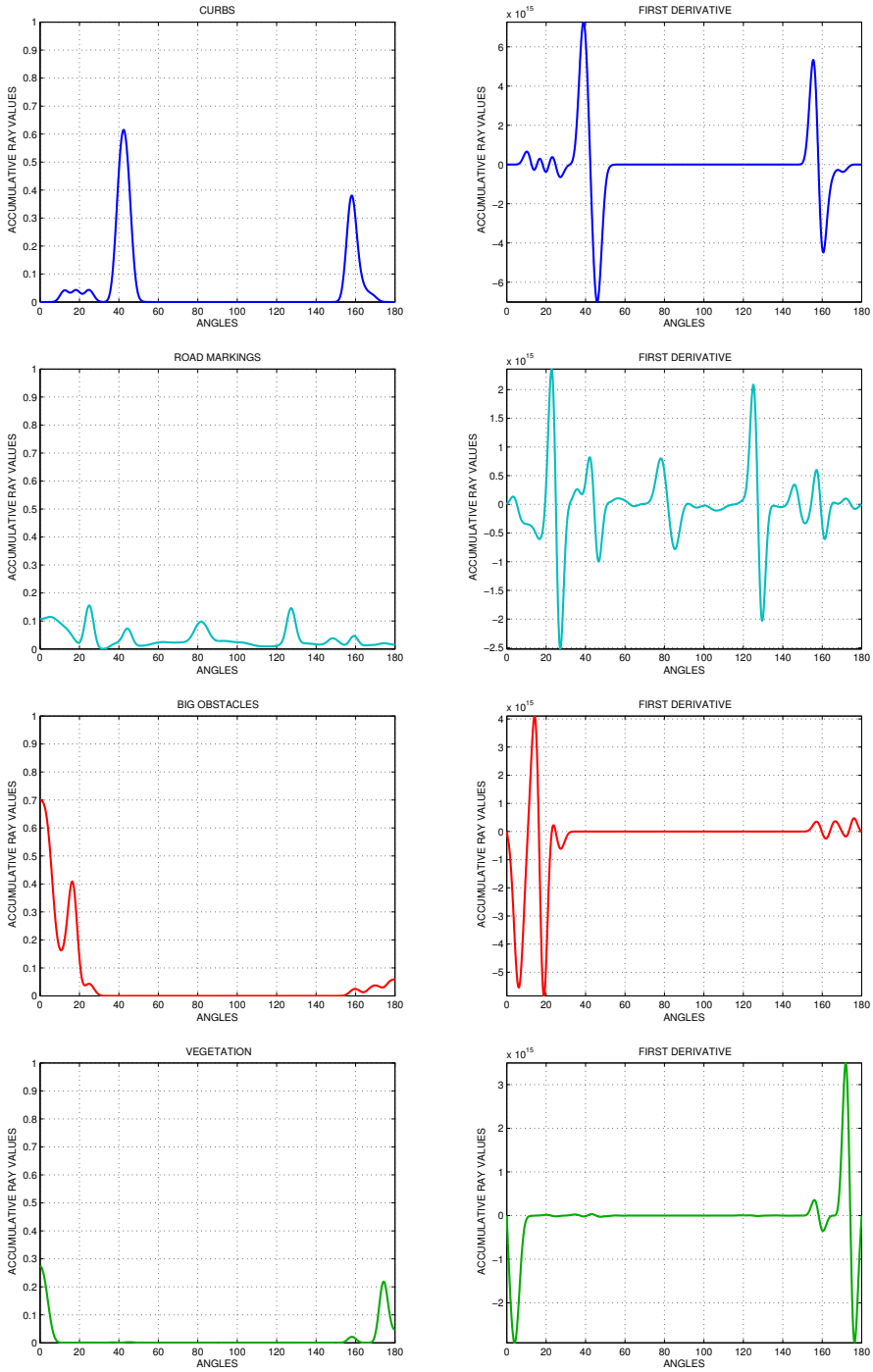


Figure 3.23: General description of the algorithm to obtain a road model.



(a) Accumulated features. (b) Features first derivative.

Figure 3.24: Graph analysis of the rays.

vegetation detector is very low, because of its robustness, instead of establishing complex conditions to the feature first derivative, a fixed threshold is applied directly to the feature. The equation 3.15 sets the conditions to the feature values to be included in the rays vector.

$$\begin{aligned}
&\lim_{x \to th_o^-} f'(x) \geq th_o \ , \forall x \in [0, 90] \\
&\lim_{x \to th_o^+} f'(x) < th_o \ , \forall x \in [0, 90] \\
&\lim_{x \to th_o^-} f'(x) < th_o \ , \forall x \in (90, 180] \\
&\lim_{x \to th_o^+} f'(x) \geq th_o \ , \forall x \in (90, 180]
\end{aligned} \tag{3.15}$$

After the creation of the rays vector, a second stage adds the lateral distance in meters of each ray to the camera origin. It is accomplished converting the ray projection into a BEV. The third step is a high level filter that makes the following assumptions:

- The ego vehicle is on the road.
- If the road has limits from the vegetation detector, the further candidates are discarded.
- If the ego vehicle is on the road, then candidates from the obstacles detector which are just in front of the ego vehicle should be removed because they are another moving vehicle instead of a free space limit.
- Some road marking candidates are arrows or other road marking symbols different from the lines that are useful to detect road limits. Dashed and solid lines have a regular pattern along the ray, contrary to that, the symbols are isolated peaks in the ray analysis. For that reason, road marking candidates that do not satisfy this condition are also removed.
- Some of the candidates are very close to each other. For example, a road marking and a curb are frequently close to each other.

In order to reduce the number candidates, both are merged in a single candidate with the mean angle of both candidates.

All the previous assumptions reduce significantly the number of candidates. That is very important during the fourth stage, where a recursive function finds all possible adjacent lanes with a specific range of valid width. In addition to the previous condition, lane widths have to be similar between each other.

The algorithm is flexible and adapts the lane range depending on the type of the road because it takes information from a digital navigation map. Digital maps include the number of lanes and the type of the road. Some of the road types are *highway*, *primary*, *secondary*, *tertiary* or *residential*. The last two types correspond to roads that usually do not have road markings, for this reason the lane width has less restrictive conditions on these roads.

The unfiltered lane combinations are shown in Figure 3.25. As in the previous step of the algorithm, the resulted road models require to satisfy some high level restrictions.

1. The ego vehicle is on the road.
2. The number of lanes should match with the information stored in the map.
3. In residential streets where there are road markings, the ego vehicle is driving on the right lane.
4. The model should have a small height difference between lanes.

The first condition removes the combinations of Figures 3.25a and 3.25c. The core of the algorithm is an iterative loop that calls to the recursive function with different sets of parameters, starting with hard restrictive conditions and relaxing them on each iteration. If at the end of the loop there is not any valid combination of lanes, the number of lanes is decremented and the loop is called again. For example, a street

(a) One lane on the left.

(b) One lane in the middle.

(c) One lane on the right.

(d) Two lanes on the left.

(e) Two lanes on the right.

(f) Three lanes.

Figure 3.25: All possible lane combinations without high level filtering.

with 2 lanes of 3 meters has not road markings and it is limited by curbs. At the first iteration, the restrictions for that road expect 2 lanes of 3 meters, however, there is not any combination that satisfies the restrictions. In the next iteration, the algorithm finds a single lane of 6 meters width, which fits to the real scene.

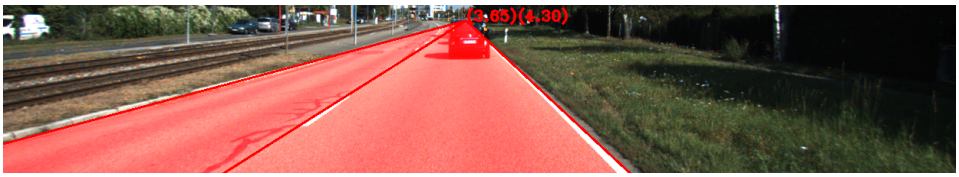
If lane markings are not correctly detected, the possibility to find a valid lane combination is reduced specially when the number of lanes is greater or equal to 3 because instead of finding 3 lanes of 3 meters each, it finds one lane of 3 meters and another one of 6 meters, which does not satisfy the condition of similar width between lanes.

Figures 3.25d and 3.25e are from a one way street, however if the street were a two way street, the third condition would remove 3.25e. In the fourth condition, the ego lane is detected (Figure 3.25b) and a plane is fitted to its 3D points using RANSAC. The combination with the smallest mean distance from the plane to the other lanes is preserved and the others are discarded.

Figures 3.26a and 3.26b show the correct road segmentation based on curbs on the left and vegetation areas on the right. Figures 3.26c and 3.26d show scenes with three lanes where the road boundaries are also composed of curbs and vegetation. Figure 3.26e shows a challenging street in the city center. Finally, in Figure 3.26f, the system fails in the scene interpretation due to the information extracted from the map and the real situation. Even though the curb is detected and the street is wide enough for two lanes, the real free space is only one lane because the other lane is occupied by parked cars.

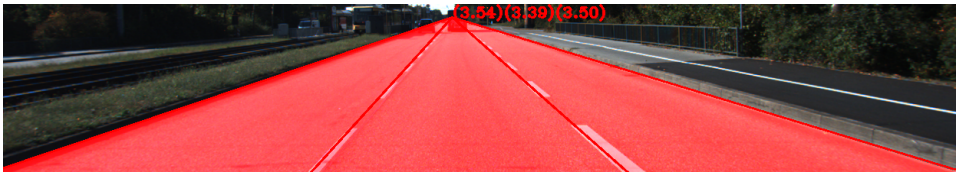
This novel approach converts important features that are not suitable to be included directly to the classifier into a road model. It reads the number of lanes and the type of road from digital navigation maps to adapt the filtering parameters for an optimal road segmentation. The detected road is estimated only from features of the stereo cameras, the color camera and the digital navigation map, which makes the system free of any type of machine learning technique.

(a) Two lanes in a *secondary* road.



(b) Two lanes in a *secondary* road.

(c) Three lanes in a *primary* road.



(d) Three lanes in a *primary* road.

(e) One lane in a city center street.

(f) Two lanes in a *residential* street with parked vehicles.

Figure 3.26: Results of road segmentation based on context information.

3.4. Road Shape Prior Obtained From Digital Maps

In section 3.3, the width of the free space is estimated using information of the road type and the number of lanes from the digital navigation map. In this section, a road prior is generated from the road width estimated in section 3.3 and the road shape provided by the digital navigation map.

This method creates a relationship between the map and the road segmentation method where both algorithms take information from each other. As detailed in Figure 3.27, the road segmentation method gets the road type and the number of lanes from the map, and it also sends the road width to the map in order to create the road prior. It is a kind of symbiosis where both functions take benefits from the other.

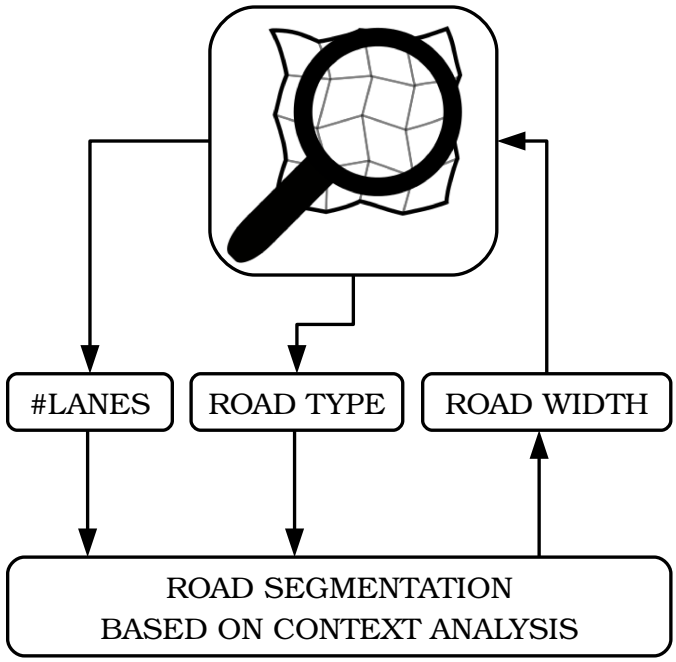
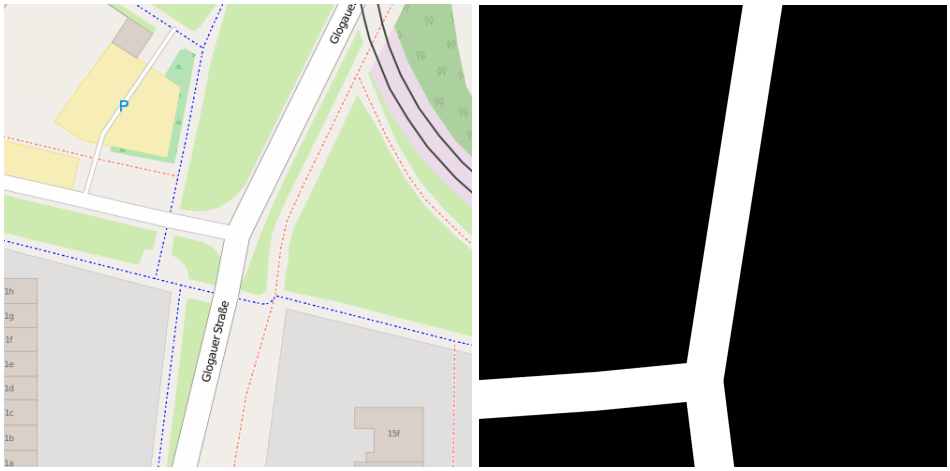


Figure 3.27: The map prior requires a road width, which is obtained from the road model presented in section 3.3. Furthermore, the generated road model uses the number of lanes and the type of the road from the map. It is a kind of symbiosis where both functions take benefits from the other.

The digital navigation maps used for our approach are Open Street Map [110]. These collaborative maps are created by a large community around the world and all the information stored in the map is editable and it is freely accessible, see Figure 3.28a. The map consists of a list of streets called ways. Every way is composed of a list of nodes with a location and its relations with the other nodes and ways. As detailed in Figure 3.28, thanks to the location and relation between the nodes, the shape of the current street and its surroundings can be estimated.



(a) Open Street Map visualization using the standard layer.

(b) Simplified representation of the intersection.

Figure 3.28: Standard layer map and line segment representation of an intersection. The orientation of the map is aligned with the vehicle orientation and the road width is estimated using the method presented in section 3.3.

The creation of a valid road prior requires to transform the map orientation to the current heading of the vehicle. The location and heading of the ego vehicle are read from a GPS/IMU sensor. They are necessary to find the current street in the map and create a simplified map with the current street and the others that are connected to it. The sensors provided in the dataset have a position accuracy of $[0.01, 1.5]$ meters depending on the satellite constellation and a heading of $0.05°$ RMS. Instead of using expensive sensors, an alternative is the use of a low cost GPS, a compass and visual odometry to improve

the accuracy of the location and the heading.

Since the road prior is created in a zenithal view, it is necessary to project the map into the image plane. The complete projection process requires to transform from the GPS coordinate system to the camera coordinate system and also to project the point into the image plane. The process is detailed in equation 3.16, where P_{gps} is the point in the GPS coordinate system, RT is the rotation and translation matrix to the camera coordinate system, R_r is the image rectification matrix and P_p is the projection matrix to the image plane. Finally, s is a scale factor and the pixel coordinates are stored in P_i.

$$s^{1\times1} \cdot P_i^{3\times1} = P_p^{3\times3} \cdot R_r^{3\times3} \cdot RT^{3\times4} \cdot P_{gps}^{4\times1} \tag{3.16}$$

The result of the road prior is shown in Figure 3.29, where different scenarios have been processed using the estimated road width and the shape obtained from the map. As demonstrated in Figures 3.29a and 3.29b, the correct localization of the vehicle creates a good prior of the road shape. The map information is specially useful in presence of intersections where the road detection is more complex, see Figures 3.29c and 3.29d. The nodes of the way are referenced to the center of the way. However, sometimes there is a drift between the map and the real center of the street. That is the case detailed in Figures 3.29e and 3.29f.

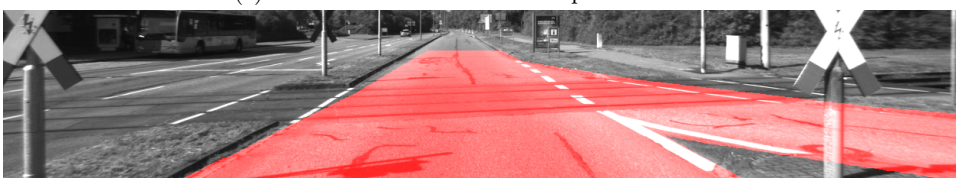
Given that the map drift is usually constant along the street, the model can be displaced along the measurements to estimate the offset and future models can be adjusted with the estimated offset. However, as the maps are updated by many collaborators, the drift can vary from one street to another, requiring to estimate the drift in every street.

In order to mitigate this offset, the final road prior is obtained modeling the uncertainty of the vehicle position and orientation with a variability of ±2 meters and ±10 degrees respectively. Furthermore, the road width is also modeled with an uncertainty of ±1 meter.

(a) Two lanes scene with correct localization.

(b) Road with multiple lanes.

(c) Road exit modeled with map information.



(d) Incoming lane in a train crossing.

(e) Road intersection with incorrect localization.

(f) Incorrect localization in a single lane street.

Figure 3.29: Results of road prior based on map shape.

(a) Road model of an adjacent street.

(b) Road model of a three an intersection of three streets.

Figure 3.30: Road prior obtained after modeling the uncertainty of the vehicle position and orientation.

Figure 3.30 shows the final road prior after modeling the uncertainty of the vehicle position and orientation. This model is used as a probability of a pixel to belong to a road pixel. This new feature is included in the feature vector of the Matching Learning (ML) classifier described in section 3.5.

3.5. Road Segmentation Based On ML

The road detection is defined in this thesis as a binary classification problem with the labels *road/non road*. The selected approach is based on Conditional Random Fields, which are extensively used for image classification problems due to their smooth and filtered result.

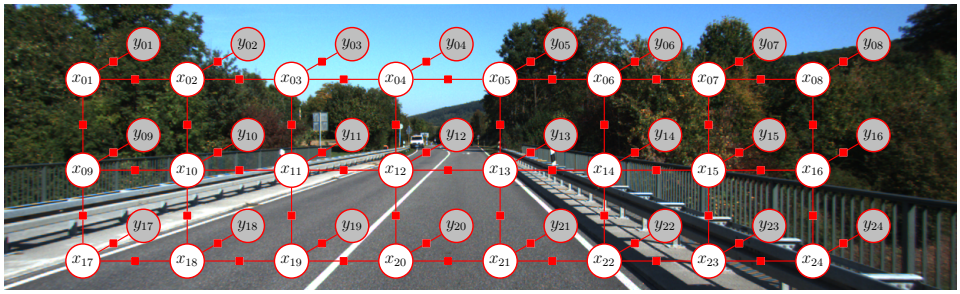


Figure 3.31: Simplified Conditional Random Field graph over the image. White nodes represent labels and grey nodes represent feature vectors.

The graph in Figure 3.31 shows the structure of the proposed CRF. It is defined as a bidimensional graph where every node represents a pixel in the image. The white nodes represent the labels assigned to the pixels and the gray ones represent the feature vectors associated to the pixels.

3.5.1. CRF

Theorem 3.5.1. *Bayes' theorem states that if x and y are events and the probability $P(y) \neq 0$, the conditional probability $P(x \mid y)$ is equal to the probability of observing event y given that x is true multiplied by the probability of observing x divided by the probability of observing y.*

$$p\left(x \mid y\right) = \frac{p\left(y \mid x\right) p\left(x\right)}{p\left(y\right)} \tag{3.17}$$

The posterior probability $p\left(x \mid y\right)$ is the probability of x given the evidence y. It contrasts with the likelihood function, which is $p\left(y \mid x\right)$. The CRFs directly model the posterior distribution $P\left(\mathbf{x} \mid \mathbf{y}\right)$ as a Gibbs field [111], therefore CRFs are discriminative models.

CRF Definition. *Let $G = (S, E)$ be a graph such that $\boldsymbol{x}$ is indexed by the vertices of G. Then $(\boldsymbol{x}, \boldsymbol{y})$ is said to be a conditional random field if, when conditioned on $\boldsymbol{y}$, the random variables x_i obey the Markov property with respect to the graph: $P\left(x_i \mid \boldsymbol{y}, \boldsymbol{x}_{\mathcal{S}-\{i\}}\right) = P\left(x_i \mid \boldsymbol{y}, \boldsymbol{x}_{\mathcal{N}_i}\right)$,*

where $S - \{i\}$ is the set of all nodes in the graph except the node i, $\mathcal{N}_i$ is the set of neighbors of the node i in G, and $\boldsymbol{x}_\Omega$ represents the set of labels at the nodes in set Ω.

In contrast to CRFs, MRFs are generative models because they describe how a label vector $\mathbf{x}$ can probabilistically "generate" a feature vector $\mathbf{y}$. One of the most common methods tries to maximize the data likelihood. The principal advantage of discriminative modeling is that it is better suited to including rich, overlapping features.

Applying CRFs to the image labeling problem, they are undirected graphical models that can be used to consider context information for the image labeling problem by modeling statistical dependencies between the labels and the data at neighboring sites [112, 113].

Given image data $\mathbf{y}$ consisting of M image sites $i \in \mathcal{S}$ with observed data y_i , i.e., $\mathbf{y} = (y_1, y_2, \ldots, y_M)^T$, where $\mathcal{S}$ is the set of all sites, we want to assign a discrete class label x_i from a given set of classes $\mathcal{C}$ to each site i. Collecting the class labels x_i in a vector $\mathbf{x} = (x_1, x_2, \ldots, x_M)^T$, we want to find the label configuration $\widehat{x}$ that maximizes the posterior probability of the labels given the data $p(\mathbf{x} \mid \mathbf{y})$, thus $\widehat{x} = arg\, max_x\, p(\mathbf{x} \mid \mathbf{y})$. CRF are discriminative models that directly model the posterior probability $p(\mathbf{x} \mid \mathbf{y})$ [112, 113]:

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{1}{Z} exp\left(\sum_{i \in \mathcal{S}} \varphi_i(x_i, \mathbf{y}) + \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j, \mathbf{y})\right) \tag{3.18}$$

In equation 3.18, Z is a normalization constant and $\mathcal{N}_i$ is the neighbourhood of data site i. The unary potential φ_i links the class label x_i of image site i to the data $\mathbf{y}$, whereas the pairwise potential $\psi_{i,j}$ models the dependencies between the labels (x_i, x_j) of neighbouring sites i, j and the data $\mathbf{y}$.

3.5.1.1. Unary Terms

Unary potential can be seen as a measure of how likely a site i will take label x_i given image features $\mathbf{y}$ ignoring the effects of other sites in the image. Boosting techniques are becoming very relevant in the road classification problem [114]. This technique combines the performance of many weak classifiers to produce a strong classifier. The weak classifier is computationally fast and it is usually a decision tree.

Instead of using decision trees as weak classifiers, they also can be used for classification, where each tree leaf is marked with a class label and multiple leaves may have the same label. Random trees is a collection of decision trees, because of that, it is also known as random forest. Every decision tree takes the input feature vector, classifies it and the forest output is the class label that received more votes. During the training stage, at each tree node, a random subset of features are used to find the best split value, in contrast, extremely randomized trees choose the feature index and the split value randomly.

Table 3.2: Comparison of tree based classifiers depending on the split value and the feature selection. The max depth is the maximum depth of each weak classifier.

CLASSIFIER	FEATURE	SPLIT VALUE	MAX DEPTH
DT	best	best	1
RT	random	best	N
ERT	random	random	N
Boosting	best	best	N

In order to find the best classifier for the road detection problem in urban scenarios, the following classifiers are compared: Boosting Discrete (BoostD), Boosting Gentle (BoostG), Extremely Randomized Trees (ERT), Random Trees (RT) and Decision Trees (DT). In Table 3.2, different tree based classifiers are compared depending on the split value and the feature selection.

Every classifier uses the feature vector detailed in Table 3.3. The length of the whole vector is 60, however in section 4.3.1, different feature combinations are evaluated to obtain their influence in the final classification result. After the training, the classifier response is normalized to the interval $[0, 1]$, where small values mean non road area and high values mean road, see Figure 3.32. In the following, this normalized value will be called unary potential.

Table 3.3: Feature vector and their feature length for the unary term classifier.

FEATURE	**LENGTH**
Normals	3
Curvatures	5
Big Obstacles	1
Vegetation	1
Heights	1
HOG	36
HSV	3
Illuminant Invariant	1
LBP	1
Pixel Location	2
XYZ	3
Shadows	1
From Limits To Road	1
Map Prior	1
TOTAL	**60**

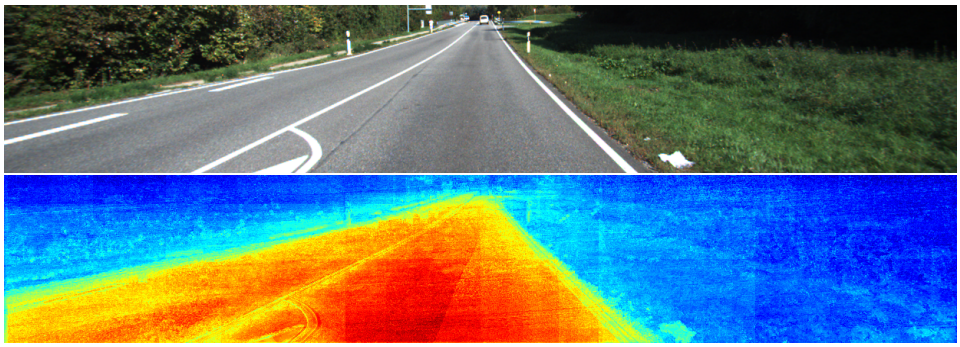


Figure 3.32: Normalized response of the classifier. The result is represented in the colorjet color scale, where cold color means non road areas and warm color means road areas.

3.5.1.2. Pairwise Terms

The pairwise term can be seen as a measure of how labels at neighboring sites i and j should interact given the observed image feature $\mathbf{y}$. The goal in the pairwise term is to have similar labels at a pairs of sites for which the observed data supports such a hypothesis. In other words, we are interested in learning a pairwise discriminative model.

The pairwise term has been trained as well as the unary term using decision trees based techniques. The pairwise term measures the relationship between the nodes and its neighbors. In order to find the best operator to relate neighbors, the following feature vectors are taken into account, where (x, y) means the node coordinates in the graph.

1. Given a feature vector $\mathbf{F}(x, y)$ of a node $n(x, y)$, and their 2 neighbors $\mathbf{F}(x-1, y)$ and $\mathbf{F}(x, y-1)$, the pairwise term is trained with the feature vector $\mathbf{F}_1(x, y) = \mathbf{F}(x, y-1) - \mathbf{F}(x, y)$ and $\mathbf{F}_1(x, y) = \mathbf{F}(x-1, y) - \mathbf{F}(x, y)$, $\forall x, y \geq 1$.

2. Given a feature vector $\mathbf{F}(x, y)$ of a node $n(x, y)$, and their 2 neighbors $\mathbf{F}(x-1, y)$ and $\mathbf{F}(x, y-1)$, the pairwise term is trained with the feature vector $\mathbf{F}_2(x, y) = |\mathbf{F}(x, y-1) - \mathbf{F}(x, y)|$ and $\mathbf{F}_2(x, y) = |\mathbf{F}(x-1, y) - \mathbf{F}(x, y)|$, $\forall x, y \geq 1$.

3. Given a unary potential $U(x, y)$ of a node $n(x, y)$, and their 4 neighbors $U(x, y-1), U(x, y+1), U(x-1, y), U(x+1, y)$, the pairwise term is trained with the feature vector $\mathbf{F}_3(x, y) = U(x, y-1) \parallel U(x, y+1) \parallel U(x-1, y) \parallel U(x+1, y)$, $\forall x, y \geq 1$.

4. Given a unary potential $U(x, y)$ of a node $n(x, y)$, and their 8 neighbors N within a radius $r = 1$, the pairwise term is trained with the Local Binary Pattern (LBP) of N. $\mathbf{F}_4(x, y) = LBP_N(x, y)$, $\forall x, y \geq 1$.

5. Given a unary potential $U(x, y)$ of a node $n(x, y)$, the pairwise potential is the same as the unary potential without any training stage $\forall x, y \geq 0$.

The feature vector of the 5 different approaches has the length detailed in Table 3.4. After the training, the classifier response is normalized to the interval $[0, 1]$, where small values mean non road area and high values mean road. In the following, this normalized value will be called pairwise potential.

Table 3.4: Feature vector and their feature length for the pairwise term classifier.

FEATURE	LENGTH
$\mathbf{F}_1$	60
$\mathbf{F}_2$	60
$\mathbf{F}_3$	4
$\mathbf{F}_4$	1

3.5.1.3. Inference

Efficient inference is critical for CRFs, both during training and for predicting the labels on new inputs. For discrete variables, the marginals could be computed by bruteforce summation, but the time required to do so is exponential in the size of $\mathbf{y}$. There are a number of exact inference algorithms for general graphical models. Although these algorithms require exponential time in the worst case, they can still be efficient for graphs that occur in practice.

The most popular exact algorithm, the junction tree algorithm, successively groups variables until the graph becomes a tree. Once an equivalent tree has been constructed, its marginals can be computed using exact inference algorithms that are specific to trees. Because of the complexity of exact inference, an enormous amount of effort has been devoted to approximate inference algorithms. Two classes of approximate inference algorithms have received the most attention: Monte Carlo [115] algorithms and variational algorithms [116]. Monte Carlo algorithms are stochastic algorithms that attempt to approximately produce a sample from the distribution of interest. Variational algorithms are algorithms that convert the inference problem into an

optimization problem, by attempting to find a simple approximation that most closely matches the intractable marginals of interest.

In our implementation, Direct Graphical Models C/C++ library [117] is used. This library has a variety of methods for learning, inference and decode tasks on Markov Random Fields (MRF) and Conditional Random Fields (CRF). The implemented inference algorithms are:

1. Inference for Markov chains based on the Chapman-Kolmogorov equations.
2. Exact inference, which is only available if $|\mathbf{x}|^{|\mathcal{S}|} < 2^{32}$.
3. Sum product Loopy Belief Propagation for lattice graphs.
4. Inference for tree undirected graphs.
5. Max product Viterbi inference algorithm for MRF.

Due to the lattice structure of our graph, see Figure 3.31, the inference has been computed using Loopy Belief Propagation [118].

3.6. Conclusion

The approach exploited in this thesis is then the collection, in a robust manner, of some features, descriptors or properties from the sensors and some priori information from digital navigation maps and then use classification techniques for the final decision whether there is road or not road, in each part of the captured images.

Particularly, the features described in this chapter embrace appearance features to describe the texture and color of the road, geometry based features to obtain the shape of the road and the geometry of the obstacles present on the road.

In addition features related with context information are analyzed in a high level interpretation of the scene. Digital navigation maps can

update map information using the context based model generated in our system.

Finally, a bidimensional CRF is proposed to filter the result of the classifier using loopy belief propagation.

Chapter 4

Results

This chapter discusses the results of the algorithm described in chapter 3 using the following structure. Section 4.1 explains the evaluation method that measures the quality of our method. Section 4.2 justifies the selection of one of the classifiers presented in section 3.5.1.1. Section 4.3.1 discusses the results obtained in the unary potentials and section 4.3.2 the ones after the CRF inference with the unary and pairwise terms.

The road detection images are compared with some of the algorithms ranked in the KITTI benchmark in section 4.4. To conclude the chapter, a discussion of some situations where our algorithm does not detect the road properly are analyzed in section 4.5.

4.1. Evaluation

The evaluation method to measure the quantitative results is the $F_1 - score$ because the score is used in an important benchmark of road detection called KITTI [119]. The dataset and benchmark is open-access and it includes 600 annotated training and test images of scenes with high variability from the KITTI autonomous driving project. The group of images is divided in three types of scenes: the

first one is urban marked road (UM), the second one is urban multiple marked lanes (UMM) and the third one is urban unmarked roads (UU).

The use of this metric and the KITTI dataset in our algorithm, make the presented method comparable with others. The score is the harmonic mean of precision and recall and it is calculated following equation 4.3.

$$precision = \frac{TP}{TP + FP} \tag{4.1}$$

$$recall = \frac{TP}{TP + FN} \tag{4.2}$$

$$F_1 - score = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{4.3}$$

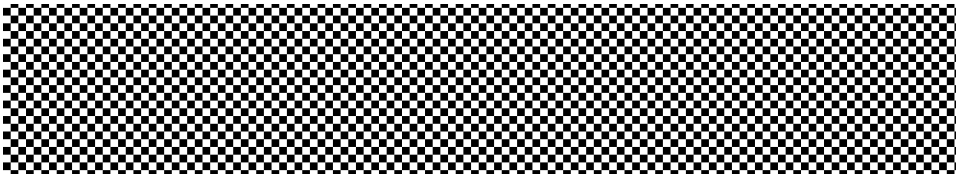
The $F_1 - score$ has been used in two different images. The first one is the image plane, which evaluate the performance in a pixel level. This is the most common approach in the literature, however in a vehicle scenario, its control stage usually happens in a 2D Bird's Eye View (BEV). The KITTI benchmark has a ranking sorted by $F_1 - score$ calculated on the BEV images. In order to compare our system with other algorithms in an international benchmark, the same evaluation method is adopted.

In the image plane, every pixel has the same weight in the global statistics, consequently, a false positive (FP) at 7 meters has the same effect as the one at 40 meters. Nevertheless, when a pixel in image plane is converted to the BEV, the further pixels get more importance in the global score.

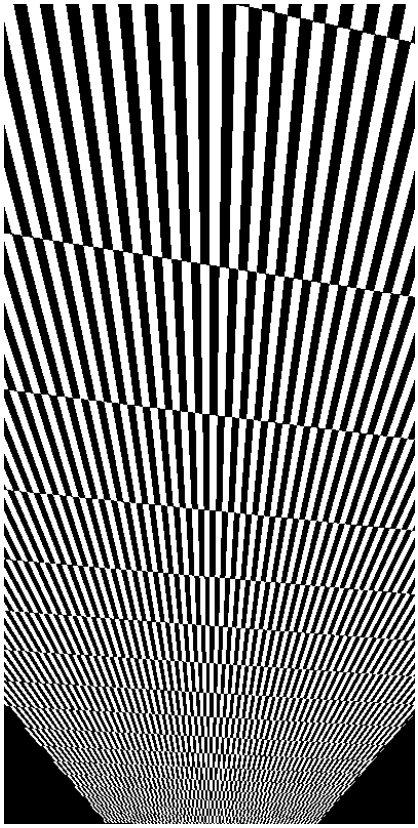
The BEV is the projection of the plane in front of the camera in a zenithal view. The first step for the correct reconstruction is to detect the ground plane. It is accomplish using stereo vision to generate the 3D points and RANSAC to fit the points to a plane. The second step is to project the ground plane in the new perspective, which is obtained using equation 4.4, where P_p is the projection matrix, R_r is

the rectification matrix, P_i is point in the image plane and s is a scale factor.

$$P_{world}^{4\times1} = \left[P_p^{3\times3} \cdot R_r^{3\times3} \cdot RT^{3\times4}\right]^{-1} \cdot P_i^{3\times1} \cdot s^{1\times1} \tag{4.4}$$



(a) Grid of squares of 10 pixel in the image plane.



(b) BEV.

Figure 4.1: Graphical demonstration of how the area of the further squares in the image plane occupy a larger area in the BEV.

Figure 4.1a shows a grid of squares of 10 pixels in the image plane.

The further pixels in the image plane occupy a larger area in the BEV than the closer ones, see Figure 4.1b. The resulted BEV is dependent of a scale factor, which is set to 20 pixels / meter. The area represented in the BEV is from 6 meters to 46 meters long and from -10 to 10 meters wide with respect to the camera coordinates.

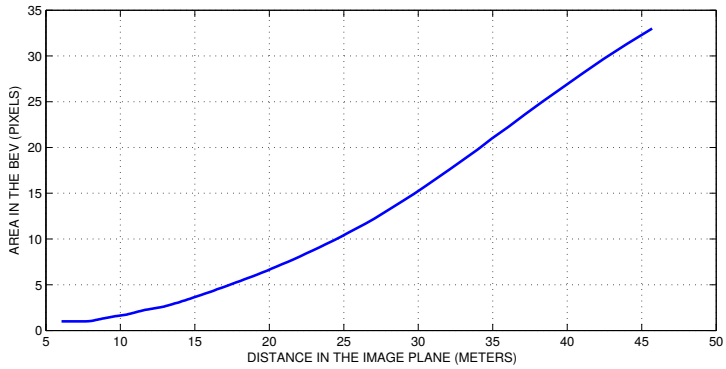


Figure 4.2: The abscissa represent the distance in meters of one single pixel in the image plane. The ordinate represent the size in pixels in the BEV of 1 pixel in the image plane.

The increment of the area is evaluated numerically in Figure 4.2. One pixel in the image plane at 17 meters occupy an area of 5 pixels in the BEV. In addition, it is remarkable that one pixel in the image plane at 40 meters occupy 27 pixels in the BEV, which is close to 12 times larger than a pixel at 11 meters. The $F_1 - score$ is also affected by this situation. Consequently, the evaluation in the BEV benefits to algorithms with high precision at long distances.

4.2. Classifier Selection

As mentioned in section 3.5.1.1, the following classifiers are compared: Boosting Discrete (BoostD), Boosting Gentle (BoostG), Extremely Randomized Trees (ERT), Random Trees (RT) and Decision Trees (DT). Given the basic feature vector detailed in Table 4.1, an analysis of how the classifier parameters affect the performance is explained in this section.

Table 4.1: Basic feature selection for the classifier parameters adjustment.

FEATURE	LENGTH	SELECTED
Normals	3	✓
Curvatures	5	✓
Big Obstacles	1	✓
Vegetation	1	✓
Heights	1	✓
HOG	36	✓
HSV	3	✓
Illuminant Invariant	1	✓
LBP	1	✓
Pixel Location	2	✓
XYZ	3	✓
Shadows	1	
From Limits To Road	1	
Map Prior	1	
TOTAL	**60**	**57**

The most important parameters to adjust are:

- **Type of classifier:** The chosen classifiers are Decision Trees (DT), Random Trees (RT), Extremely Randomized Trees (ERT), Discrete AdaBoost (BoostD) and Gentle AdaBoost (BoostG).
- **Number of weak classifiers:** The analyzed values for this parameter are 50, 100, 250 and 500.
- **Maximum Depth:** The maximum depth of each weak classifier. The analyzed values for this parameter are 5, 10 and 25.
- **Cost of a missclasification for a specific label:** If w_{NR} is 1 and w_R is 10, then each mistake in predicting the label *road* is equivalent to making 10 mistakes in predicting the label *non road*. The analyzed values for this parameter are 1 and 10.

The classifiers have been trained with the same number of samples ($\sim 4.5M$) and features (57). The most important aspects to choose the best classifier are the memory requirements and the performance.

The selection of the best classifier is decided in three steps: The first step evaluates the performance. Given Figure 4.3a, the classifiers with the greatest performance are Gentle Boosting and Boosting Discrete, regardless of the selected tree depth.

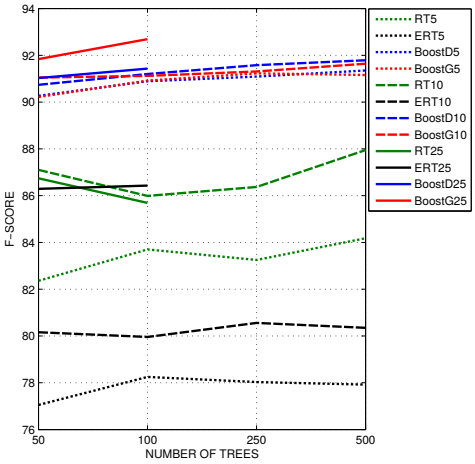
The second step evaluate the memory usage. Figure 4.3b shows the memory requirements for all the classifiers. It is observed that boosting classifiers with depths greater than 5 require high amounts of memory. In our system we assume a maximum of ~ 1 GB for the road classifier, which discard classifiers with more than 250 trees and depths greater than 5.

The trade off between performance and memory requirements is represented in Figure 4.3c, where the classifier Gentle Boost with 250 trees and a depth of 5 is the one with the best balance between memory usage and performance. The results discussed in the rest of the document are obtained using that classifier.

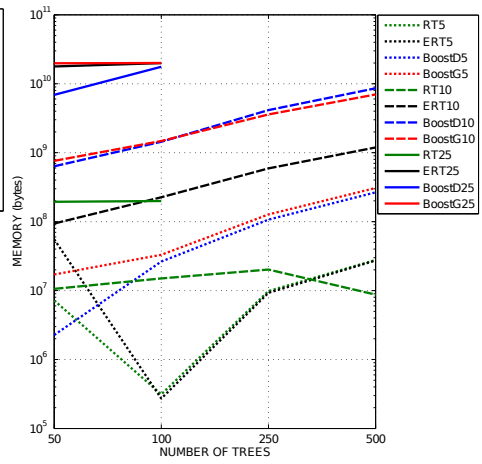
4.3. CRF

As explained in 3.5.1, the two main terms in the CRF are the unary and the pairwise terms. The most important one is the unary term because is strongly related with the final response of the CRF. The pairwise term smooths the unary response and filters some missdetection and false positive values.

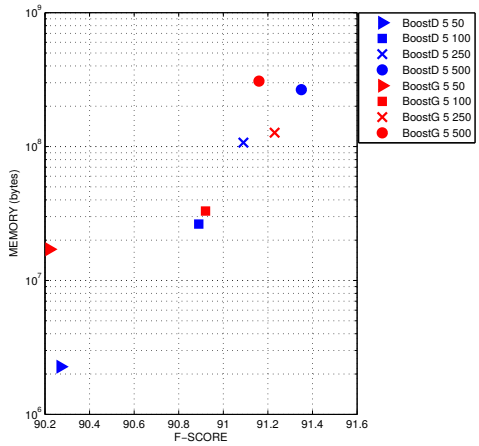
In section 4.2, a deep analysis of different classifiers is performed. The result of the trade off between performance and memory requirements outcomes that the Gentle Boosting classifier with a maximum of 250 trees of depth of 5 is the best option to train and classify road pixels. In this section, different features are taken into account to evaluate the influence of each one in the final response of the unary and pairwise terms.



(a) F-score calculated in the image plane.



(b) Memory usage during training.



(c) Performance vs memory usage.

Figure 4.3: Selection of the best classifier.

4.3.1. Unary Terms

As in section 4.2, the classifier has been trained with 50% of the available training data and the other 50% is used for testing. The number of samples per image is in most of the images $\sim 465K$ pixels. The amount of samples for the 150 images is intractable ($\sim 69M$), therefore a subsampling technique is applied to reduce the number of samples up to $\sim 30K$ per image. The first step is to remove samples above the row 155, which is over the horizon line. The second one is to

take 1/3 of the pixels in the horizontal and vertical dimensions. Finally the amount of samples for the training stage is reduced to ($\sim 4.5M$).

Given a basic feature set composed of texture, color, shape and 3D information, three more features have been tested to evaluate their influence in the final response. In Table 4.2, $F_1 - score$ is computed in perspective and BEV images on UM, UMM and UU scenes.

Table 4.2: Performance comparison of unary potentials trained with different combination of features.

	Image plane				Bird Eye View			
Feature Set	**UM**	**UMM**	**UU**	**All**	**UM**	**UMM**	**UU**	**All**
Basic	91.25	89.20	85.92	88.73	86.35	85.88	77.07	82.99
Basic + Shadow	91.43	89.32	86.19	88.92	86.83	86.40	77.17	83.35
Basic + Context	92.76	90.44	88.25	90.43	89.02	86.03	79.09	84.60
Basic + Map	91.49	92.03	88.42	90.61	87.76	89.30	80.49	85.76
Full Set	92.59	93.17	89.69	91.78	88.94	89.86	81.36	86.63

In average, the shadow detector increases 0.36% the performance compared with the basic feature set. The improvement is similar in all the scenes. In spite of that, the road segmentation based on context features, increases the score specially in scenes with two lanes with road markings (UM). The improvement of 2.67% and 2.02% in UM and UU scenes respectively contrast with the 0.15% of the UMM scenarios.

As explained in section 3.3, the road detector based on context information, adds the lateral distance for each limit and tries to find a combination of lanes that satisfies the restrictions of number of lanes and similar lane width. In scenarios like UMM, most of the images has 2, 3 or more lanes. If road markings are not correctly detected in every single lane, the function could generate an inaccurate model of the road.

For example, if the road has 3 lanes of 3 meter width and the system detects one lane of 3 meters and another one of 6 meters, the lanes combination will be discarded because of the difference between lane

widths. That is the reason for the imperceptible improvement in UMM scenes.

The map prior obtained from the digital navigation map provides an important information of the road shape. This feature increases the performance 3.42% in UMM and UU scenarios and 1.41% in UM with respect to the basic feature set.

The join feature set of *basic + shadow + context + map prior* obtains a score of 88.94% in UM scenes, which is 0.08% below the combination *basic + context*. Nevertheless, in average the full set of features increases the performance 3.64%.

Some qualitative and quantitative results are shown in the image plane and BEV in Figures 4.4, 4.5 and 4.6 on UM, UMM and UU scenes respectively. In Figure 4.4, the combination *basic + context* fits very well to the real shape of the road in UM. Consequently, when the map prior is included to the feature set, the classifier interprets part of the prior information as noise or unreliable data in the final response.

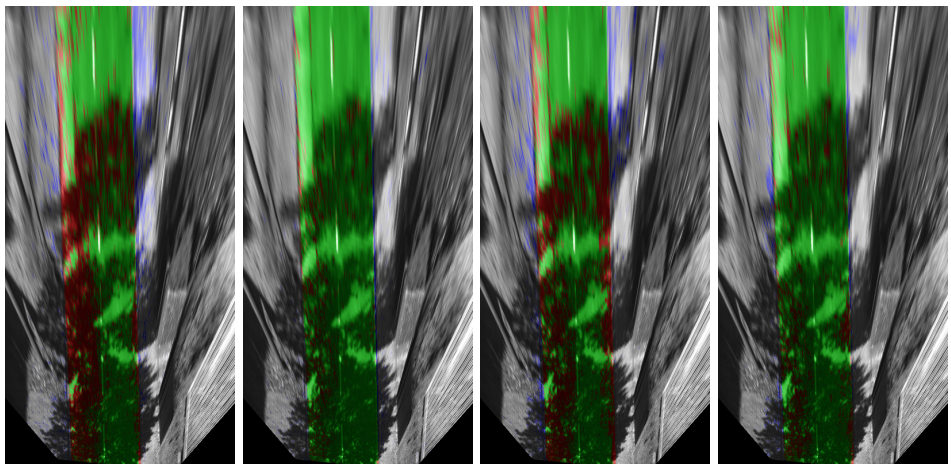
In scenes with multiple marked lanes, see Figure 4.5, the *map prior* is very important because sometimes the *context feature* does not provide reliable information.

Finally, in Figure 4.6 all the features combinations have good results. However, the combination *basic + context* obtains less false positives, specially at far distances, which is very important in the BEV evaluation.

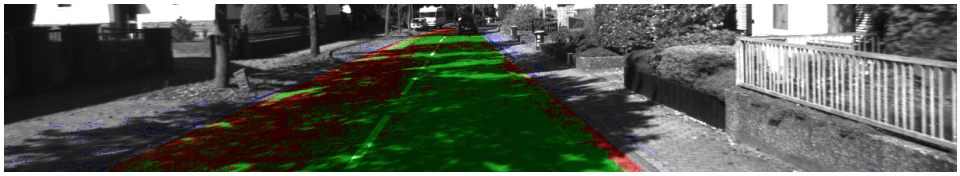
4.3.2. Pairwise Terms

As in section 4.2, the classifier has been trained with 50% of the available training data and the other 50% is used for testing.

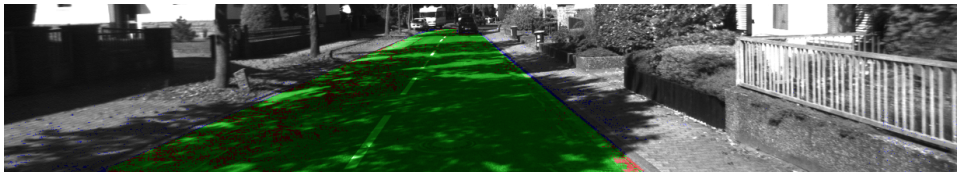
A subsampling technique is also applied to reduce the amount of samples for the training stage to ($\sim 4.5M$), however the feature vector size differs from one test to another, see Table 4.3. Four different feature vectors have been tested:



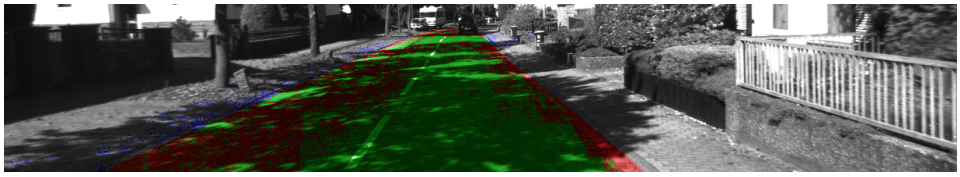
(a) Basic (b) Basic + Context (c) Basic + Map (d) Full feature set



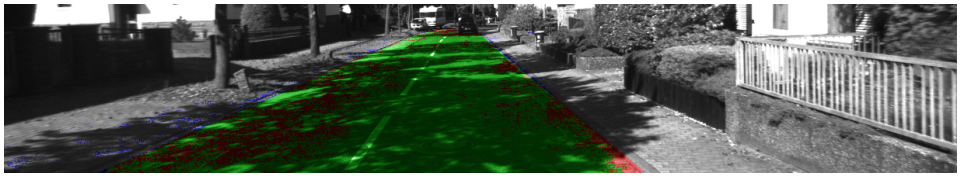
(e) Basic



(f) Basic + Context



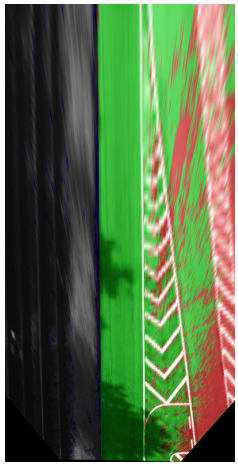
(g) Basic + Map



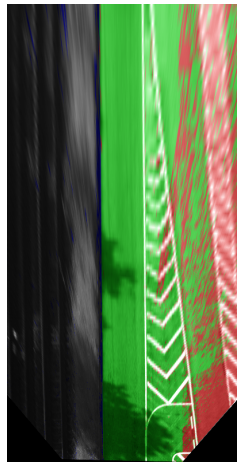
(h) Full feature set

FEATURE SET	BASIC	BASIC + CONTEXT	BASIC + MAP	FULL SET
BEV	73.01	94.36	77.01	91.40
Image Plane	77.86	93.91	78.07	88.22

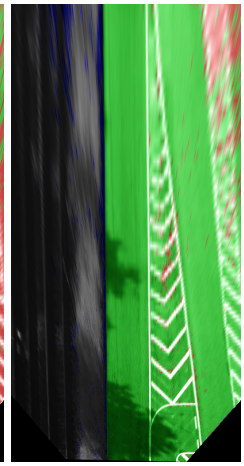
Figure 4.4: Results of unary terms in a specific UM image.



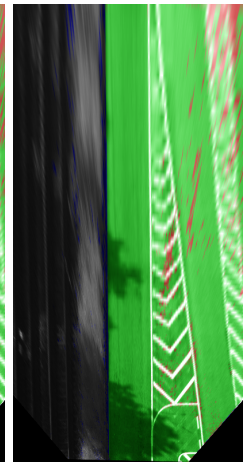
(a) Basic



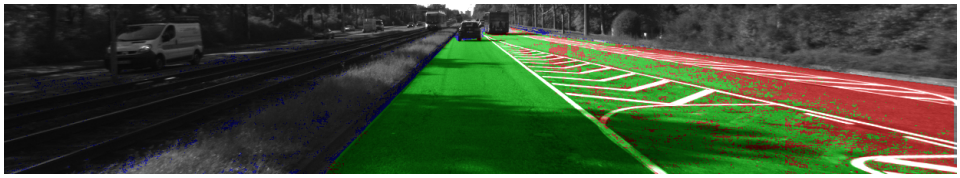
(b) Basic + Context



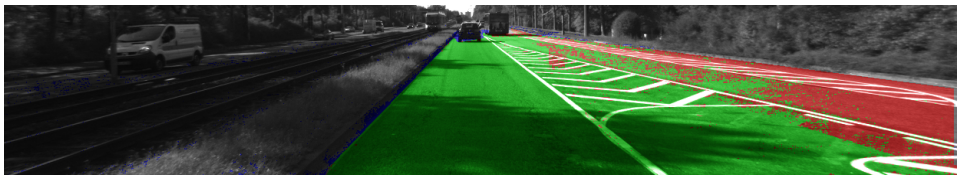
(c) Basic + Map



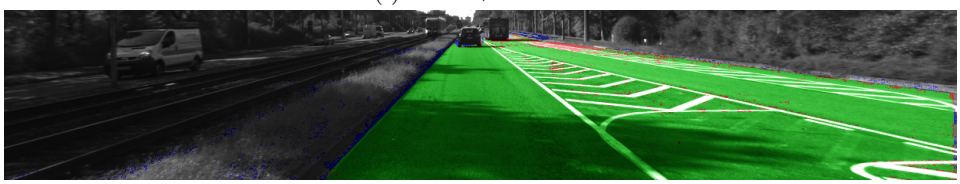
(d) Full feature set



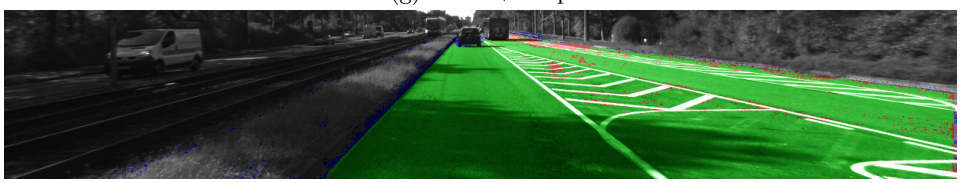
(e) Basic



(f) Basic + Context



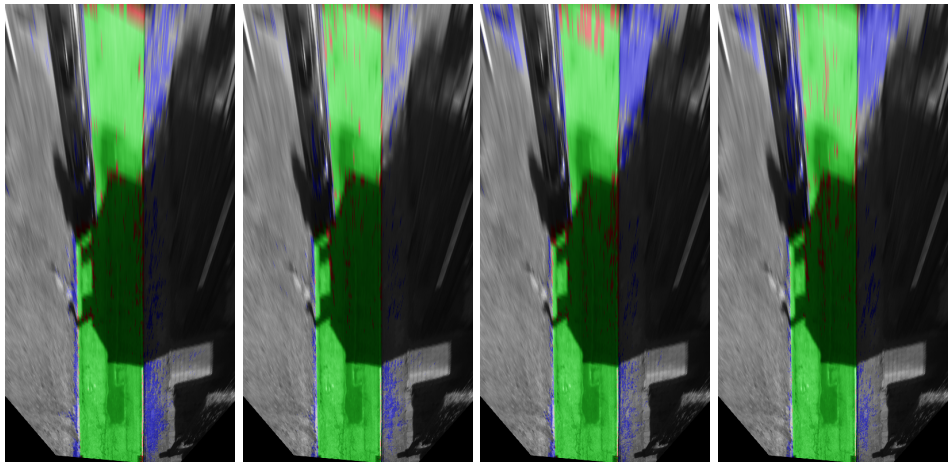
(g) Basic + Map



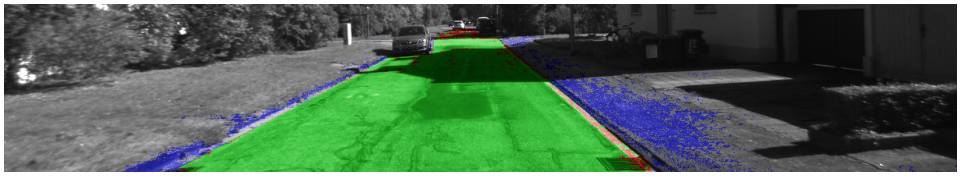
(h) Full feature set

FEATURE SET	BASIC	Basic + Context	BASIC + MAP	FULL SET
BEV	79.64	81.63	94.12	95.58
Image Plane	81.21	85.52	97.06	96.87

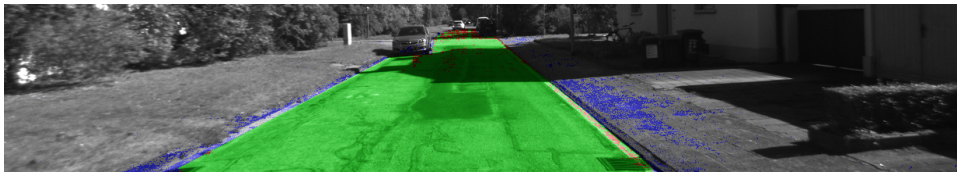
Figure 4.5: Results of unary terms in a specific UMM image.



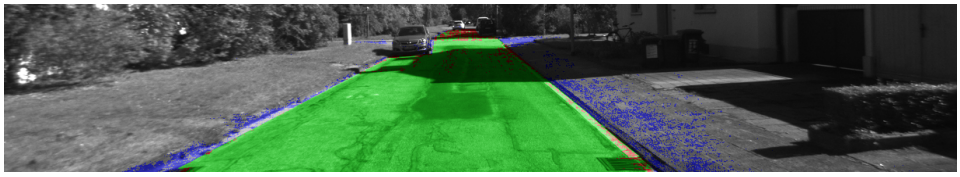
(a) Basic (b) Basic + Context (c) Basic + Map (d) Full feature set



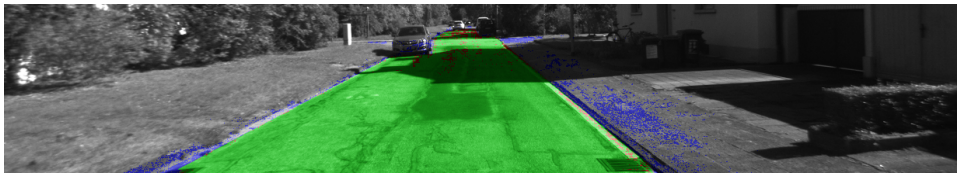
(e) Basic



(f) Basic + Context



(g) Basic + Map



(h) Full feature set

FEATURE SET	BASIC	Basic + Context	BASIC + MAP	FULL SET
BEV	88.30	91.94	85.04	87.09
Image Plane	88.62	93.84	92.61	94.15

Figure 4.6: Results of unary terms in a specific UU image.

1. Given a feature vector $\mathbf{F}(x, y)$ of a node $n(x, y)$, and their 2 neighbors $\mathbf{F}(x-1, y)$ and $\mathbf{F}(x, y-1)$, the pairwise term is trained with the feature vector $\mathbf{F}_1(x, y) = \mathbf{F}(x, y-1) - \mathbf{F}(x, y)$ and $\mathbf{F}_1(x, y) = \mathbf{F}(x-1, y) - \mathbf{F}(x, y)$, $\forall x, y \geq 1$.

2. Given a feature vector $\mathbf{F}(x, y)$ of a node $n(x, y)$, and their 2 neighbors $\mathbf{F}(x-1, y)$ and $\mathbf{F}(x, y-1)$, the pairwise term is trained with the feature vector $\mathbf{F}_2(x, y) = |\mathbf{F}(x, y-1) - \mathbf{F}(x, y)|$ and $\mathbf{F}_2(x, y) = |\mathbf{F}(x-1, y) - \mathbf{F}(x, y)|$, $\forall x, y \geq 1$.

3. Given a unary potential $U(x, y)$ of a node $n(x, y)$, and their 4 neighbors $U(x, y-1), U(x, y+1), U(x-1, y), U(x+1, y)$, the pairwise term is trained with the feature vector $\mathbf{F}_3(x, y) = U(x, y-1) \parallel U(x, y+1) \parallel U(x-1, y) \parallel U(x+1, y)$, $\forall x, y \geq 1$.

4. Given a unary potential $U(x, y)$ of a node $n(x, y)$, and their 8 neighbors N within a radius $r = 1$, the pairwise term is trained with the Local Binary Pattern (LBP) of N. $\mathbf{F}_4(x, y) = LBP_N(x, y)$, $\forall x, y \geq 1$.

5. Given a unary potential $U(x, y)$ of a node $n(x, y)$, the pairwise potential is the same as the unary potential without any training stage $\forall x, y \geq 0$.

Table 4.3: Feature vector and their feature length for the pairwise term classifier.

FEATURE	LENGTH	SAMPLES
$\mathbf{F}_1$	60	$\sim 9.0M$
$\mathbf{F}_2$	60	$\sim 9.0M$
$\mathbf{F}_3$	4	$\sim 4.5M$
$\mathbf{F}_4$	1	$\sim 4.5M$

The pairwise terms are not tested individually, nevertheless, they are validated jointly with the unaries and the inference method. Before the inference, a basic pre-processing filter is applied to the unary response. The filter finds individual contours and removes all the contours except the contour with the largest area, which is the road. In

addition, the holes found into the road contour are filled to obtain a consistent road shape.

Table 4.4: Performance comparison of the CRF output using unary potentials and different pairwise potentials.

	Image Plane				Bird Eye View			
Pairwise	**UM**	**UMM**	**UU**	**All**	**UM**	**UMM**	**UU**	**All**
Feature Difference	93.34	94.04	85.40	90.83	**89.01**	**92.09**	83.00	**87.96**
Absolute Feature Diff.	92.57	92.93	83.91	89.70	88.04	92.02	80.55	86.76
Neighbors Potentials	92.44	93.71	89.51	91.85	86.89	88.89	82.06	85.88
LBP of Unaries	92.28	93.54	89.33	91.68	86.60	88.59	81.61	85.54
Unaries	91.97	94.90	91.40	92.74	88.76	90.83	**84.10**	87.84

Table 4.4 shows the results of the CRF with different pairwise potentials, where difference of features is the pairwise term with the best average results. It is remarkable how the unary potentials applied to the pairwise obtain an average performance similar to the difference of features.

Table 4.5 confirm the importance of the unary potential to the final response of the CRF. However, the smoother result of the CRF and the filtering process improves the performance 1.33%. Some graphical results of the road detection using CRF are shown in Figures 4.7, 4.8 and 4.9.

Table 4.5: Performance comparison of the unary potentials and the CRF using feature difference as pairwise term.

	Image Plane				Bird Eye View			
Method	**UM**	**UMM**	**UU**	**All**	**UM**	**UMM**	**UU**	**All**
With CRF (Feature Difference)	93.34	94.04	85.40	90.83	89.01	92.09	83.00	87.96
Without CRF (Only Unary)	92.59	93.17	89.69	91.78	88.94	89.86	81.36	86.63

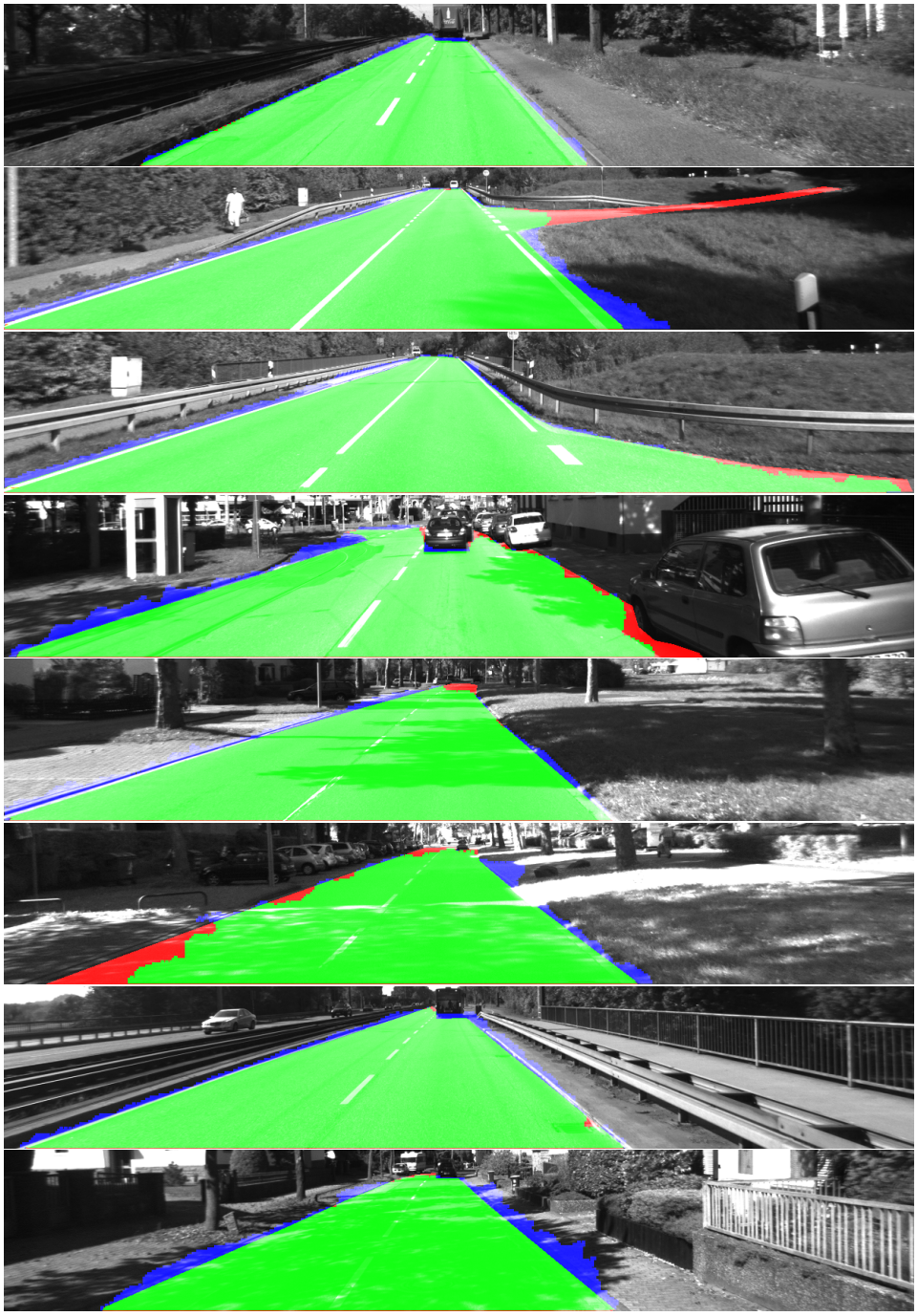


Figure 4.7: Final road detection results using a CRF in UM scenes.

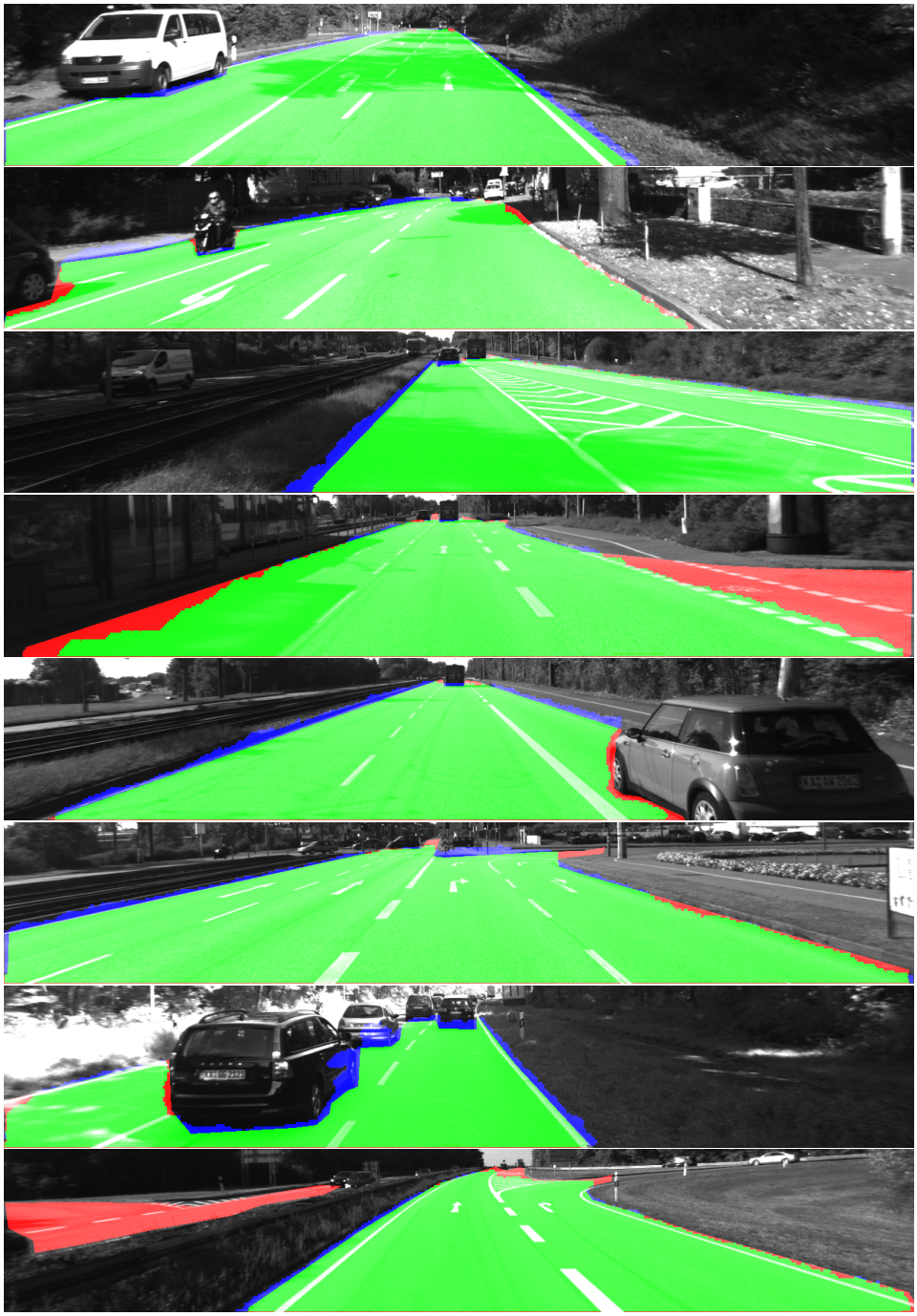


Figure 4.8: Final road detection results using a CRF in UMM scenes.

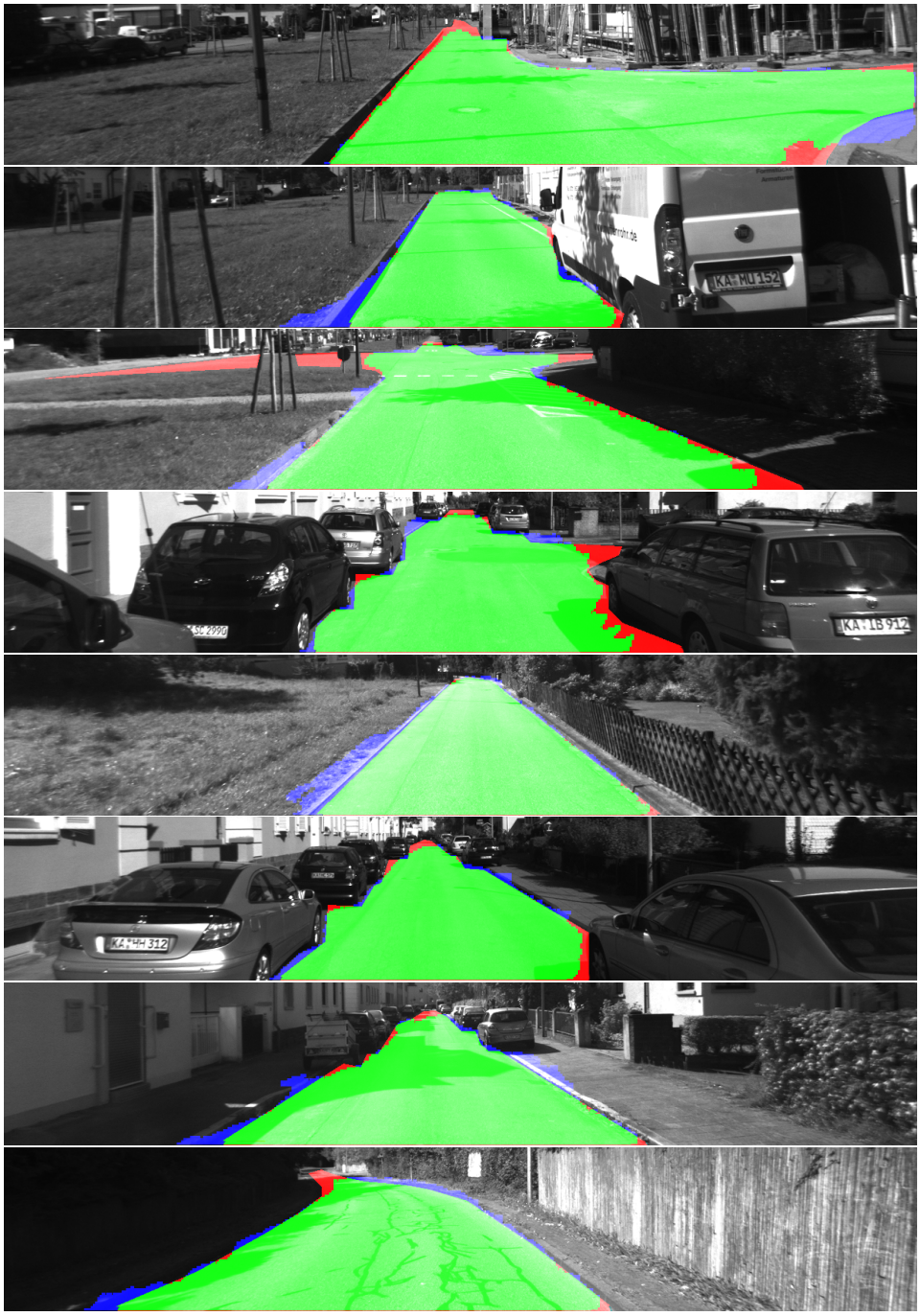


Figure 4.9: Final road detection results using a CRF in UU scenes.

4.4. Comparative Results

As mentioned in previous sections, the road detection method described in this thesis is developed and tested in the KITTI benchmark and dataset. As shown in Table 4.6, the first positions of the KITTI ranking are occupied by Convolutional Neural Networks (CNN) based methods. These new approaches have been a revolution to the semantic segmentation problem since 2015.

Table 4.6: Performance comparison of our method with the algorithms of the KITTI benchmark.

METHOD	DESCRIPTION	UM	UMM	UU	All
SSL[94]	CNN	93.94	96.01	93.19	94.38
HIM[120]	RF+hierarchical graphical model	90.07	93.55	85.76	89.79
NNP[95]	NN+CRF	90.50	91.34	85.55	89.14
CB[91]	contextual blocks	88.89	90.55	86.13	88.52
Ours	**Boosting+CRF**	**89.01**	**92.09**	**83.00**	**87.96**
SPRAY[92]	spatial ray features	88.14	89.69	82.71	86.84
ProbBoost[121]	joint boosting	87.48	91.36	80.76	86.53
BL	avg(ground truth)	82.24	76.02	69.50	75.92

The best method that is not based on CNN and uses vision sensors to detect the road obtains a $F_1 - score$ of 89.79%, which is only 1.83% over our proposal. Analyzing the different types of scenes, our method is in third position on UMM scenes with a score of 92.09%.

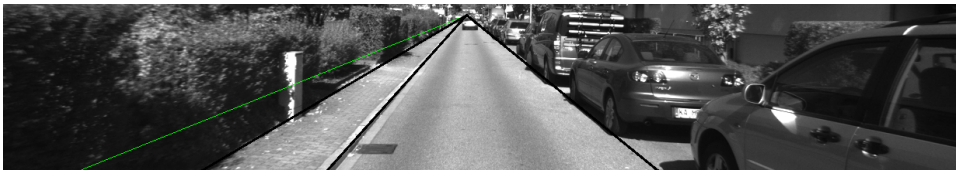
Due to urban unmarked streets are more challenging than roads with road markings, the obtained score is 83%. Some specific situations are deeply analyzed in section 4.5, where the different inputs of our system are analyzed independently.

In spite of the difficult scenes present in urban unmarked scenarios, the comparative analysis rank our new method into the best classifiers in the state of the art.

4.5. Discussion

The graphical results of Figures 4.7, 4.8 and 4.9 show scenes where the system detects correctly the drivable area. However, the system fails in other challenging situations.

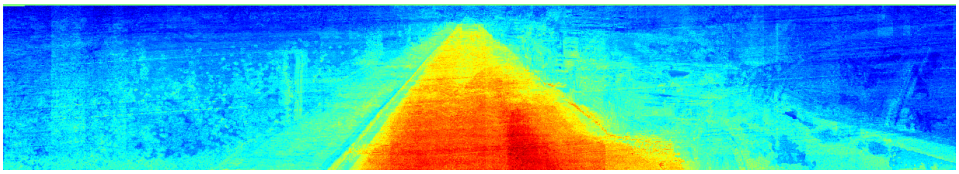
Figure 4.10 shows a residential scenario where there are two lanes but one of them is occupy by parked cars. Even though the curb



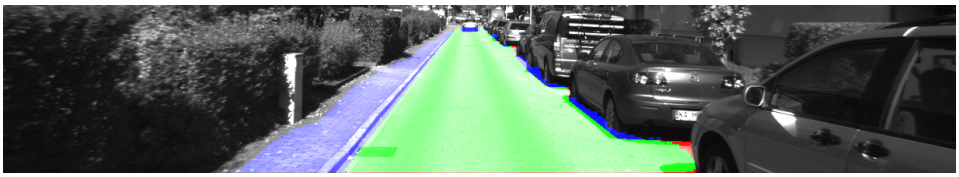
(a) Road limits.

(b) Road model based on context information.

(c) Prior based on map information.



(d) Result of the boosting classifier.



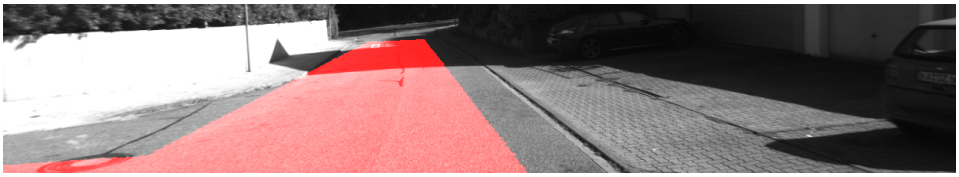
(e) Final road detection after CRF.

Figure 4.10: Challenging scenario

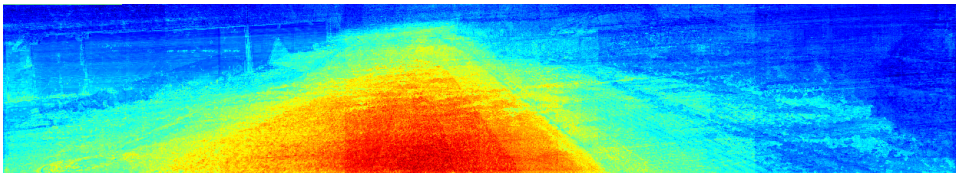
is correctly detected, the number of lanes extracted from the map does not match with the real free space and the road model based on context information generate an unrealistic model. The boosting classifier detects the road correctly but there are some pixels (yellow) on the side walk classified as road. The CRF fills the space between them to smooth the final result, which includes the side walk as drivable area.

(a) Road limits.

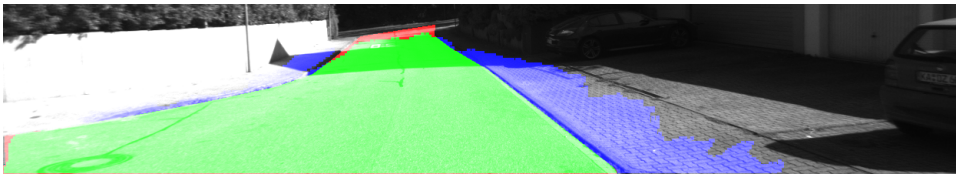
(b) Road model based on context information (not available).



(c) Prior based on map information.



(d) Result of the boosting classifier.



(e) Final road detection after CRF.

Figure 4.11: Challenging scenario

In the intersection of Figure 4.11, the road curbs are not detected neither the road marking nearby. In absence of context information the default road width value of the map creates an additional feature to the classifier, which jointly with the other basic features is able to detect the road. The classifier result has some FP beyond the curb, which are still considered as road after the CRF. However, the FP of the street on the left are converted in TP. This scene demonstrate that context information is very important to obtain a high level model of the road and feed the boosting classifier with robust features.

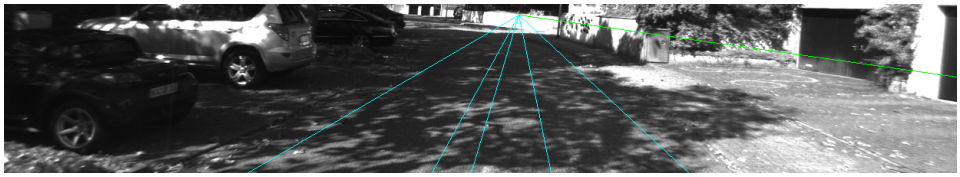
In Figure 4.12 the curbs are not detected and many road markings are detected due to the shadows created by trees. Consequently, the road model generated using context information does not fit to the real scene. The preciseness of the model is partially compensated with the prior based on the map. In spite of the shadows and the false positives of the boosting classifier in the parking area and the side walk, the CRF obtains a fairs detection of the road.

4.6. Conclusion

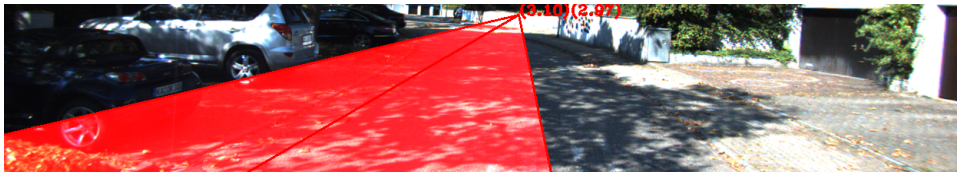
The boosting classifier has demonstrate its ability to detect the road correctly integrating types of features. Despite the CRF is a good method to obtain a smooth result, in some situations it integrates disperse FP pixels and create a large area of FP.

The BEV evaluation benefits to algorithm with high precision at long distances because of the area increment of further pixels in the BEV perspective.

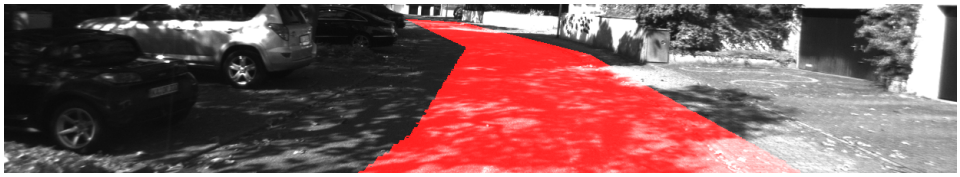
After the analysis of scenarios where the detection of the free space is complex, the conclusion is that all the components involve in the system play an important role in the final decision. Some incorrect feature values can be filtered with other features, but if the low level features fail, in most of the cases the high level features dependent of



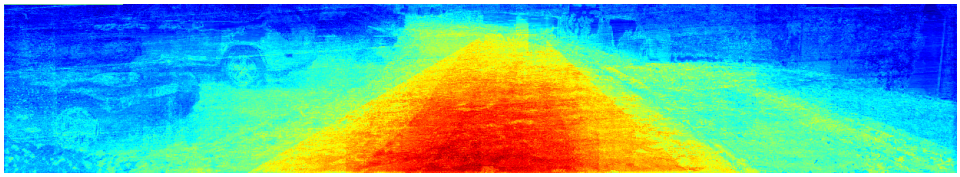
(a) Road limits.



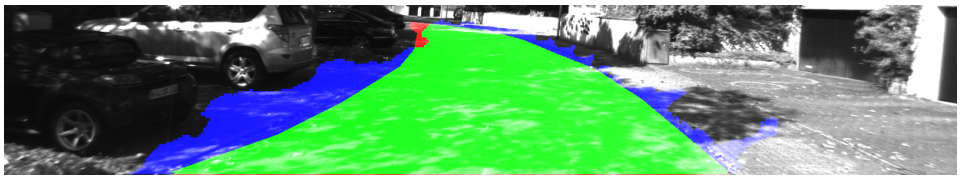
(b) Road model based on context information.



(c) Prior based on map information.



(d) Result of the boosting classifier.



(e) Final road detection after CRF.

Figure 4.12: Challenging scenario

them will fail too. Context interpretation is very important because it can model the road in very complex situations with few high level features. The incorrect interpretation of the context can generate a wrong model when the features are correctly detected. For this reason, the effort should be put in the improvement of this topic.

Chapter 5

Conclusions

This chapter presents the global conclusions and discuss the main contributions introduced and developed along the chapters of this thesis.

5.1. Curb detection

In this thesis, a novel method to detect free space is presented. The main contributions have taken place in the field of new features development. Urban scenarios present challenging situations where the road is limited by small curbs or different pavement textures. Their detection is a key point of the free space estimation for autonomous vehicles. In the curb detection task, different approaches are based on LIDAR sensors, which are more expensive than stereo cameras. For this reason and because stereo vision has a dense disparity map, our proposal is based on stereo vision.

The original method based on curvature estimation is presented in section 3.2.3.3. Other state of the art methods based on stereo vision, detect curbs of 5 cm height up to 10 meters, however, the proposed method does not require any parameter adjustment and it is able to detect a wide range of curbs. The minimum curb height required for

a precise detection is around 3 cm and the detection distance is up to 20 meters whenever the curbs are connected in the curvature image, otherwise a small object of 3 cm at 20 meters far will be filtered.

5.2. Road model based on context features

Machine learning techniques are widely applied to solve semantic segmentation problems. The proposed method uses a boosting classifier trained with a feature set that describes the road. Some of the features mentioned before, such as curbs and road markings provide important information about the road but they cannot be included directly to the road classifier.

For this reason a novel method is presented to convert from features that describe road limits to a new feature that describe road areas. Instead of creating a set of radial rays from the bottom of the image, as other methods in the literature do, our method presents a new approach that uses the vanishing point to create a set of radial rays that fits to the road limits. This new approach improves road classifier performance specially in urban marked scenes.

5.3. Road prior based on navigation map

Another contribution of this thesis is the creation of a new way to update digital navigation maps. The innovation presented in section 3.4 aims to update information of road width. The system takes the number of lanes and the road type from the digital navigation map and returns the road width. The map can be updated from several vehicles, creating a robust value of the road width. The map with the road width is used to generate a prior of the current structure of the road, which is very useful in intersections and narrow streets.

5.4. Other contributions

Our proposal takes advantage of machine learning techniques, however, the method that creates a road model based on context features obtains good results without any machine learning technique. The image processing algorithm is good enough to outline the road when the system should be applied to a new city or country scenario and the classifier has not been trained with the new images.

Chapter 6

Future Work

From the results and conclusions of the present work, several future lines for each treated topic are devised. They correspond to aspects that have not been solved or that need a further analysis to improve the performance of the system.

- The future work is focused on computation time optimization to make the system work in real time. This issue is not treated in this thesis because it is focus on the algorithm. Despite the processing time is around 30 seconds per frame, many of the image processing algorithms can be parallelized and computed on GPU. The next step is to run the road detection method into our autonomous vehicle.
- In order to take advantage of on board sensors, a sensor fusion should be implemented. Multi-sensor approaches are very important to have redundancy. The autonomous vehicle includes RADAR, LIDAR, GPS, stereo vision and color camera sensors. Each sensor is strong in some situations and weak in others, therefore, the redundancy aims to obtain a robust system which is able to work in a high variety of situations.
- Convolutional Neural Networks (CNN) outperform the state of the art in semantic segmentation problems, however they should

be trained with similar scenes. Our system will be integrated with a CNN classifier to increase its robustness in situations where the CNN has not been trained.

- Context interpretation has demonstrate an important role in the road detection problem. For this reason, an special effort will be dedicated to the improvement of this high level analysis.
- The boosting classifier is trained with features in the image plane and at the end of the process, the system is evaluated in a BEV perspective. In order to optimize the results in the evaluation perspective, a new classifier will be trained with the same features but transformed into a BEV perspective before the training stage.
- The road is very similar between consecutive frames. The detected road in the previous frame will be integrated as another new feature to the presented feature vector and the whole system will the evaluated with the new feature.

Bibliography

[1] "Global status report on road safety," World Health Organization, Tech. Rep., 2015.

[2] "Analytical report on road safety," European Commission, Tech. Rep., 2010.

[3] J. B. Consulting, "Car telephone use and road safety," An overview prepared for the European Commission, Tech. Rep., 2009.

[4] "Evaluation of regulation 443/2009 and 510/2011 on reduction of co2 emissions from light-duty vehicles," A study for the European Commission, Tech. Rep., 2015.

[5] "Road departure safety, fatal analysis reporting system (fars)," Federal Highway Administration, U.S. Department of Transportation, Tech. Rep., 2013.

[6] J. B. Greenblatt and S. Saxena, "Autonomous taxis could greatly reduce greenhouse-gas emissions of us light-duty vehicles," vol. 5, no. 1, pp. 860–863, March 2015.

[7] S. Rosenbloom, "Driving cessation among older people: When does it happen and what impact does it have?" *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1779, pp. 93–99, 2001.

[8] "Audi study of traffic light information online," http://goo.gl/7nt9eH.

[9] "The economic and environment impact of traffic congestion in europe and the us: 2013-2030," Inrix, driving intelligence, Tech. Rep., 2014.

[10] J. B. Greenblatt and S. Shaheen, "Automated vehicles, on-demand mobility, and environmental impacts," *Current Sustainable/Renewable Energy Reports*, vol. 2, no. 3, pp. 74–81, 2015. [Online]. Available: http://dx.doi.org/10.1007/s40518-015-0038-5

[11] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strauss, C. Stiller, T. Dang, U. Franke, N. Appenrodt, C. G. Keller, E. Kaus, R. G. Herrtwich, C. Rabe, D. Pfeiffer, F. Lindner, F. Stein, F. Erbs, M. Enzweiler, C. Knoppel, J. Hipp, M. Haueis, M. Trepte, C. Brenk, A. Tamke, M. Ghanaat, M. Braun, A. Joos, H. Fritz, H. Mock, M. Hein, and E. Zeeb, "Making bertha drive. an autonomous journey on a historic route," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 2, pp. 8–20, Summer 2014.

[12] P. Czerwionka, "A Three Dimensional Map Format for Autonomous Vehicles," Master's thesis, Intelligent Systems and Robotics, Freie Universität Berlin, Germany, 2014.

[13] E. Guizzo, "How google's self-driving car works," *IEEE Spectrum*, 2011. [Online]. Available: http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works

[14] Y. Yu, J. Li, H. Guan, F. Jia, and C. Wang, "Learning hierarchical features for automated extraction of road markings from 3-d mobile lidar point clouds," *IEEE Journal of Selected Topics*

in Applied Earth Observations and Remote Sensing, vol. 8, no. 2, pp. 709–726, Feb 2015.

[15] T. Li and D. Zhidong, "A new 3d lidar-based lane markings recognition approach," in *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2013, pp. 2197–2202.

[16] D. Habermann, A. Hata, D. Wolf, and F. S. Osorio, "3d point clouds segmentation for autonomous ground vehicle," in *2013 III Brazilian Symposium on Computing Systems Engineering*, Dec 2013, pp. 143–148.

[17] R. Fernandes, C. Premebida, P. Peixoto, D. Wolf, and U. Nunes, "Road detection using high resolution lidar," in *2014 IEEE Vehicle Power and Propulsion Conference (VPPC)*, Oct 2014, pp. 1–6.

[18] G. Zhao and J. Yuan, "Curb detection and tracking using 3d-lidar scanner," in *2012 19th IEEE International Conference on Image Processing*, Sept 2012, pp. 437–440.

[19] W. Yao, Z. Deng, and L. Zhou, "Road curb detection using 3d lidar and integral laser points for intelligent vehicles," in *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on*, Nov 2012, pp. 100–105.

[20] J. Liu, H. Liang, and Z. Wang, "A framework for detecting road curb on-line under various road conditions," in *Robotics and Biomimetics (ROBIO), 2014 IEEE International Conference on*, Dec 2014, pp. 297–302.

[21] T. Chen, B. Dai, D. Liu, J. Song, and Z. Liu, "Velodyne-based curb detection up to 50 meters away," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, June 2015, pp. 241–248.

[22] Y. Zhang, J. Wang, X. Wang, C. Li, and L. Wang, "A real-time curb detection and tracking method for ugvs by using a 3d-lidar sensor," in *2015 IEEE Conference on Control Applications (CCA)*, Sept 2015, pp. 1020–1025.

[23] J. Choi, S. Ulbrich, B. Lichte, and M. Maurer, "Multi-target tracking using a 3d-lidar sensor for autonomous vehicles," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, Oct 2013, pp. 881–886.

[24] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, "3d traffic scene understanding from movable platforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, pp. 1012–1025, May 2014.

[25] K. Takagi, K. Morikawa, T. Ogawa, and M. Saburi, "Road environment recognition using on-vehicle lidar," in *2006 IEEE Intelligent Vehicles Symposium*, 2006, pp. 120–125.

[26] H. Kong, J. Y. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2211–2220, Aug 2010.

[27] L. Mioulet, T. P. Breckon, A. Mouton, H. Liang, and T. Morie, "Gabor features for real-time road environment classification," in *Industrial Technology (ICIT), 2013 IEEE International Conference on*, Feb 2013, pp. 1117–1121.

[28] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, June 2005, pp. 886–893 vol. 1.

[29] J. Zhang and H. H. Nagel, "Texture-based segmentation of road images," in *Intelligent Vehicles '94 Symposium, Proceedings of the*, Oct 1994, pp. 260–265.

[30] C. Fernandez, R. Izquierdo, D. F. Llorca, and M. A. Sotelo, "A comparative analysis of decision trees based classifiers for road detection in urban environments," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Sept 2015, pp. 719–724.

[31] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, June 2010.

[32] D. K. Savitha and S. Rakshit, "Gaussian mixture model based road signature classification for robot navigation," in *Emerging Trends in Robotics and Communication Technologies (INTERACT), 2010 International Conference on*, Dec 2010, pp. 230–233.

[33] K. Lu, J. Li, X. An, and H. He, "A hierarchical approach for road detection," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 517–522.

[34] T. C. Dong-Si, D. Guo, C. H. Yan, and S. H. Ong, "Robust extraction of shady roads for vision-based ugv navigation," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2008, pp. 3140–3145.

[35] G. D. Finlayson and G. Schaefer, "Hue that is invariant to brightness and gamma," in *Proceedings of the British Machine Vision Conference 2001, BMVC 2001, Manchester, UK, 10-13 September 2001*, 2001, pp. 1–10. [Online]. Available: http://dx.doi.org/10.5244/C.15.32

[36] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 59–68, Jan 2006.

[37] J. M. Alvarez and A. M. Lopez, "Road detection based on illuminant invariance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, March 2011.

[38] J. C. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7, no. 1, pp. 20–37, 2006.

[39] S. Nedevschi, R. Schmidt, T. Graf, R. Danescu, D. Frentiu, T. Marita, F. Oniga, and C. Pocol, "3d lane detection system based on stereovision," in *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, Oct 2004, pp. 161–166.

[40] A. Wedel, H. Badino, C. Rabe, H. Loose, U. Franke, and D. Cremers, "B-spline modeling of road surfaces with an application to free-space estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 4, pp. 572–583, Dec 2009.

[41] H. Loose and U. Franke, "B-spline-based road model for 3d lane recognition," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, Sept 2010, pp. 91–98.

[42] L. Bentabet, S. Jodouin, D. Ziou, and J. Vaillancourt, "Road vectors update using sar imagery: a snake-based method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1785–1803, Aug 2003.

[43] J. Beck and C. Stiller, "Non-parametric lane estimation in urban environments," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, June 2014, pp. 43–48.

[44] A. V. Nefian and G. R. Bradski, "Detection of drivable corridors for off-road autonomous navigation," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 3025–3028.

[45] A. Broggi and S. Cattani, "An agent based evolutionary approach to path detection for off-road vehicle guidance," *Pattern Recognition Letters*, vol. 27, no. 11, pp. 1164–1173, 2006.

[46] A. Parajuli, M. Celenk, H. B. Riley *et al.*, "Robust lane detection in shadows and low illumination conditions using local gradient features," *Open Journal of Applied Sciences*, vol. 3, no. 01, p. 68, 2013.

[47] P. Foucher, Y. Sebsadji, J.-P. Tarel, P. Charbonnier, and P. Nicolle, "Detection and recognition of urban road markings using images," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on.* IEEE, 2011, pp. 1747–1752.

[48] J. Huang, H. Liang, Z. Wang, T. Mei, and Y. Song, "Robust lane marking detection under different road conditions," in *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2013, pp. 1753–1758.

[49] A. Hata and D. Wolf, "Road marking detection using lidar reflective intensity data and its application to vehicle localization," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct 2014, pp. 584–589.

[50] H. Guan, J. Li, Y. Yu, Z. Ji, and C. Wang, "Using mobile lidar data for rapidly updating road markings," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2457–2466, Oct 2015.

[51] H. Guan, J. Li, Y. Yu, and C. Wang, "Rapid update of road surface databases using mobile lidar: Road-markings," in *Geo-Information Technologies for Natural Disaster Management (GiT4NDM), 2013 Fifth International Conference on*, Oct 2013, pp. 124–129.

[52] C. Fernandez, M. Gavilan, D. F. Llorca, I. Parra, R. Quintero, A. G. Lorente, L. Vlacic, and M. A. Sotelo, "Free space and speed humps detection using lidar and vision for urban autonomous navigation," in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, June 2012, pp. 698–703.

[53] W. S. Wijesoma, K. R. S. Kodagoda, and A. P. Balasuriya, "Road-boundary detection and tracking using ladar sensing," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 3, pp. 456–464, June 2004.

[54] J. Han, D. Kim, M. Lee, and M. Sunwoo, "Enhanced road boundary and obstacle detection using a downward-looking lidar sensor," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 3, pp. 971–985, March 2012.

[55] A. Y. Hata, F. S. Osorio, and D. F. Wolf, "Robust curb detection and vehicle localization in urban environments," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, June 2014, pp. 1257–1262.

[56] Y. Zhang, J. Wang, X. Wang, C. Li, and L. Wang, "3d lidar-based intersection recognition and road boundary detection method for unmanned ground vehicle," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Sept 2015, pp. 499–504.

[57] J. Siegemund, D. Pfeiffer, U. Franke, and W. Förstner, "Curb reconstruction using conditional random fields," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, June 2010, pp. 203–210.

[58] J. Siegemund, U. Franke, and W. Förstner, "A temporal filter approach for detection and reconstruction of curbs and road surfaces based on conditional random fields," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, June 2011, pp. 637–642.

[59] F. Oniga, S. Nedevschi, and M. M. Meinecke, "Curb detection based on a multi-frame persistence map for urban driving scenarios," in *2008 11th International IEEE Conference on Intelligent Transportation Systems*, Oct 2008, pp. 67–72.

[60] F. Oniga and S. Nedevschi, "Polynomial curb detection based on dense stereovision for driving assistance," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, Sept 2010, pp. 1110–1115.

[61] ——, "Curb detection for driving assistance systems: A cubic spline-based approach," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, June 2011, pp. 945–950.

[62] W. S. Wijesoma, K. R. S. Kodagoda, A. P. Balasuriya, and E. K. Teoh, "Road edge and lane boundary detection using laser and vision," in *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, vol. 3, 2001, pp. 1440–1445 vol.3.

[63] C. Yu and D. Zhang, "Road curbs detection based on laser radar," in *2006 8th international Conference on Signal Processing*, vol. 4, 2006.

[64] B. Fardi, H. Weigel, G. Wanielik, and K. Takagi, "Road border recognition using fir images and lidar signal processing," in *2007 IEEE Intelligent Vehicles Symposium*, June 2007, pp. 1278–1283.

[65] C. Fernandez, D. F. Llorca, C. Stiller, and M. A. Sotelo, "Curvature-based curb detection method in urban environments using stereo and laser," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, June 2015, pp. 579–584.

[66] C. Rasmussen, "Combining laser range, color, and texture cues for autonomous road following," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 4, 2002, pp. 4320–4325 vol.4.

[67] F. Homm, N. Kaempchen, J. Ota, and D. Burschka, "Efficient occupancy grid computation on the gpu with lidar and radar for road boundary detection," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, June 2010, pp. 1006–1013.

[68] J. Gunnarsson, L. Svensson, L. Danielsson, and F. Bengtsson, "Tracking vehicles using radar detections," in *2007 IEEE Intelligent Vehicles Symposium*, June 2007, pp. 296–302.

[69] X. Dai, A. Kummert, S. B. Park, and U. Iurgel, "Vehicle centroid estimation based on radar multiple detections," in *Vehicular Electronics and Safety, 2007. ICVES. IEEE International Conference on*, Dec 2007, pp. 1–5.

[70] C. Lundquist, L. Hammarstrand, and F. Gustafsson, "Road intensity based mapping using radar measurements with a probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1397–1408, April 2011.

[71] F. Sarholz, J. Mehnert, J. Klappstein, J. Dickmann, and B. Radig, "Evaluation of different approaches for road course estimation using imaging radar," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 4587–4592.

[72] K. Y. Guo, E. G. Hoare, D. Jasteh, X. Q. Sheng, and M. Gashinova, "Road edge recognition using the stripe hough transform from millimeter-wave radar images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 825–833, April 2015.

[73] M. Nikolova and A. Hero, "Segmentation of a road from a vehicle-mounted radar and accuracy of the estimation," in *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, 2000, pp. 284–289.

[74] X. Liu, Z. Sun, and H. He, "On-road vehicle detection fusing radar and vision," in *Vehicular Electronics and Safety (ICVES), 2011 IEEE International Conference on*, July 2011, pp. 150–154.

[75] Y. Fang, I. Masaki, and B. Horn, "Depth-based target segmentation for intelligent vehicles: fusion of radar and binocular stereo," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 3, pp. 196–202, Sep 2002.

[76] E. Richter, R. Schubert, and G. Wanielik, "Radar and vision based data fusion - advanced filtering techniques for a multi object vehicle tracking system," in *Intelligent Vehicles Symposium, 2008 IEEE*, June 2008, pp. 120–125.

[77] U. Kadow, G. Schneider, and A. Vukotich, "Radar-vision based vehicle recognition with evolutionary optimized and boosted features," in *2007 IEEE Intelligent Vehicles Symposium*, June 2007, pp. 749–754.

[78] F. Janda, S. Pangerl, E. Lang, and E. Fuchs, "Road boundary detection for run-off road prevention based on the fusion of video and radar," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, June 2013, pp. 1173–1178.

[79] G. Alessandretti, A. Broggi, and P. Cerri, "Vehicle and guard rail detection using radar and vision data fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 95–105, March 2007.

[80] D. Zhang, B. Fang, W. Yang, X. Luo, and Y. Tang, "Robust inverse perspective mapping based on vanishing point," in *Security, Pattern Analysis, and Cybernetics (SPAC), 2014 International Conference on*, Oct 2014, pp. 458–463.

[81] S. Vacek, C. Schimmel, and R. Dillmann, "Road-marking analysis for autonomous vehicle guidance." in *EMCR*, 2007.

[82] B. Mathibela, P. Newman, and I. Posner, "Reading the road: Road marking classification and interpretation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2072–2081, Aug 2015.

[83] M. Nieto and L. Salgado, "Simultaneous estimation of vanishing points and their converging lines using the em algorithm," *Pattern Recogn. Lett.*, vol. 32, no. 14, pp. 1691–1700, Oct 2011.

[84] J. Son, H. Yoo, S. Kim, and K. Sohn, "Real-time illumination invariant lane detection for lane departure warning system," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1816–1824, 2015.

[85] Y. Wang, E. K. Teoh, and D. Shen, "Lane detection and tracking using b-snake," *Image and Vision computing*, vol. 22, no. 4, pp. 269–280, 2004.

[86] F. Moosmann and C. Stiller, "Velodyne slam," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, June 2011, pp. 393–398.

[87] R. Valencia, E. H. Teniente, E. Trulls, and J. Andrade-Cetto, "3d mapping for urban service robots," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2009, pp. 3076–3081.

[88] P. Pfaff, R. Triebel, C. Stachniss, P. Lamon, W. Burgard, and R. Siegwart, "Towards mapping of cities," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 4807–4813.

[89] J. M. Alvarez, F. Lumbreras, T. Gevers, and A. M. López, "Geographic information for vision-based road detection," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, June 2010, pp. 621–626.

[90] S. Zhou, J. Gong, G. Xiong, H. Chen, and K. Iagnemma, "Road detection using support vector machine based on online learning

and evaluation," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, June 2010, pp. 256–261.

[91] C. C. T. Mendes, V. Frémont, and D. F. Wolf, "Vision-based road detection using contextual blocks," *arXiv preprint arXiv:1509.01122*, 2015.

[92] G. B. Vitor, A. C. Victorino, and J. V. Ferreira, "A probabilistic distribution approach for the classification of urban roads in complex environments," in *Workshop on IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[93] ——, "Comprehensive performance analysis of road detection algorithms using the common urban kitti-road benchmark," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*. IEEE, 2014, pp. 19–24.

[94] R. Mohan, "Deep deconvolutional networks for scene parsing," *arXiv preprint arXiv:1411.4101*, 2014.

[95] X. Chen, K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma, S. Fidler, and R. Urtasun, "3d object proposals for accurate object class detection," in *Advances in Neural Information Processing Systems*, 2015, pp. 424–432.

[96] C. C. T. Mendes, V. Frémont, and D. F. Wolf, "Exploiting fully convolutional neural networks for fast road detection," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 3174–3179.

[97] C.-A. Brust, S. Sickert, M. Simon, E. Rodner, and J. Denzler, "Convolutional patch networks with spatial prior for road detection and urban scene understanding," *arXiv preprint arXiv:1502.06344*, 2015.

[98] L. Xiao, B. Dai, D. Liu, T. Hu, and T. Wu, "Crf based road detection with multi-sensor fusion," in *Intelligent Vehicles Symposium (IV), 2015 IEEE*. IEEE, 2015, pp. 192–198.

[99] M. Passani, J. J. Yebes, and L. M. Bergasa, "Fast pixelwise road inference based on uniformly reweighted belief propagation," in *Intelligent Vehicles Symposium (IV), 2015 IEEE*. IEEE, 2015, pp. 519–524.

[100] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[101] J. M. Alvarez, A. Lopez, and R. Baldrich, "Illuminant-invariant model-based road segmentation," in *Intelligent Vehicles Symposium, 2008 IEEE*, June 2008, pp. 1175–1180.

[102] G. Finlayson and M. Drew, "4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2, 2001, pp. 473–480 vol.2.

[103] G. D. Finlayson, S. D. Hordley, and M. S. Drew, "Removing shadows from images," in *Proceedings of the 7th European Conference on Computer Vision-Part IV*, ser. ECCV '02. London, UK, UK: Springer-Verlag, 2002, pp. 823–836. [Online]. Available: http://dl.acm.org/citation.cfm?id=645318.649239

[104] G. D. Finlayson, M. S. Drew, and C. Lu, *Computer Vision - ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part III*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, ch. Intrinsic Images by Entropy Minimization, pp. 582–595.

[105] R. B. Rusu, "Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments," Ph.D. dissertation, Technische Universität München, 2009.

[106] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," in *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '92. New York, NY, USA: ACM, 1992, pp. 71–78.

[107] E. Shaffer and M. Garland, "Efficient adaptive simplification of massive meshes," in *Proceedings of the Conference on Visualization '01*, ser. VIS '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 127–134.

[108] M. Nieto, "Detection and tracking of vanishing points in dynamic environments." Ph.D. dissertation, Universidad Politécnica de Madrid (UPM), 2010.

[109] S. Choi, T. Kim, and W. Yu, "Performance evaluation of RANSAC family," in *British Machine Vision Conference, BMVC 2009, London, UK, September 7-10, 2009. Proceedings*, 2009, pp. 1–12.

[110] "Official website of open street map (osm) project." http://www.openstreetmap.org.

[111] E. Behrends, *Introduction to Markov Chains: With Special Emphasis on Rapid Mixing*. Wiesbaden: Vieweg+Teubner Verlag, 2000, ch. Markov random fields, pp. 183–194. [Online]. Available: http://dx.doi.org/10.1007/978-3-322-90157-6_19

[112] S. Kosov, F. Rottensteiner, and C. Heipke, "3d classification of crossroads from multiple aerial images using conditional random fields," in *Pattern Recognition in Remote Sensing (PRRS), 2012 IAPR Workshop on*, Nov 2012, pp. 1–4.

[113] S. Kumar and M. Hebert, "Discriminative random fields," *International Journal of Computer Vision*, vol. 68, no. 2, pp. 179–201, 2006.

[114] A. Torralba, K. P. Murphy, and W. T. Freeman, "Contextual models for object detection using boosted random fields," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005, pp. 1401–1408.

[115] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

[116] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, no. 1-2, pp. 1–305, jan 2008. [Online]. Available: http://dx.doi.org/10.1561/2200000001

[117] S. Kosov, "Direct graphical models c++ library," http://research.project-10.de/dgm, 2013.

[118] P. F. Felzenszwalb and D. R. Huttenlocher, "Efficient belief propagation for early vision," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, June 2004, pp. I–261–I–268 Vol.1.

[119] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[120] D. Munoz, J. A. Bagnell, and M. Hebert, "Stacked hierarchical labeling," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 57–70.

[121] T. Kühnl, F. Kummert, and J. Fritsch, "Spatial ray features for real-time ego-lane extraction," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. IEEE, 2012, pp. 288–293.